

AN ANCIENT DNA STUDY OF FOUR SYMPATRIC
SPECIES OF MOA (AVES: DINORNITHIFORMES) FROM
HOLOCENE DEPOSITS IN NORTH CANTERBURY,
SOUTH ISLAND, NEW ZEALAND

A thesis submitted in partial fulfilment of the requirements for the Degree
of Doctor of Philosophy in Ecology
in the University of Canterbury

by Morten Erik Allentoft

University of Canterbury

2010

Contents

Acknowledgements	5
Abstract	6
Chapter 1: Introduction	7
Eighteen years of ancient DNA research on extinct New Zealand moa (Aves: Dinornithiformes)	8
1.1.1 Introduction	8
1.1.2 Moa evolution and phylogeography	10
1.1.3 Taxonomy and genetic identification	15
1.1.4 Moa biology	18
1.1.5 The future of moa genetics	19
1.1.6 References	23
Aims and scopes of the thesis	27
1.2.1 Aims and scopes	27
1.2.2 References	29
The North Canterbury moa	31
1.3.1 <i>Dinornis robustus</i> (South Island giant moa)	32
1.3.2 <i>Euryapteryx curtus</i> (Stout-legged moa)	32
1.3.3 <i>Pachyornis elephantopus</i> (Heavy-footed moa)	32
1.3.3 <i>Emeus crassus</i> (Eastern moa)	33
1.3.4 The North Canterbury environment	33
1.3.5 The sites	35
1.3.6 References	37
Chapter 2: Basic genetic components: taxon, sex and preservation	39
Introducing the manuscripts in Chapter 2	40
Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand's extinct megafauna	41
2.1.1 Abstract	41
2.1.2 Introduction	42
2.1.3 Regional setting	44
2.1.4 Materials and methods	46
2.1.4.1 Sampling	46
2.1.4.2 DNA extraction, species determination, and qPCR	46
2.1.4.3 Molecular sexing	47
2.1.5 Results	48
2.1.5.1 Species representation	48
2.1.5.2 Molecular sexing	49
2.1.6 Discussion	52
2.1.6.1 Accuracy of morphological taxonomic identification	52
2.1.6.2 Overall species abundances	53
2.1.6.3 Niche separation, biased deposition or unrepresentative collections?	53
2.1.6.4 Sex ratios	55
2.1.7 Conclusions	58
2.1.8 Acknowledgements	58
2.1.9 References	59
Degradation of ancient DNA in bone – insights from three megafaunal sites in New Zealand	63
2.2.1 Abstract	63

2.2.2 Introduction	63
2.2.3 Materials and methods	66
2.2.3.1 Discovery and description of the site	66
2.2.3.2 Sampling	67
2.2.3.3 ¹⁴ C AMS dating	68
2.2.3.4 Ancient DNA extraction, PCR and preservation	68
2.2.4 Results	70
2.2.4.1 Morphological description	70
2.2.4.2 ¹⁴ C AMS ages	71
2.2.4.3 Genetic species determination	72
2.2.4.4 qPCR results and nuclear DNA preservation	72
2.2.4.5 DNA preservation through time	73
2.2.5 Discussion	74
2.2.5.1 Age, species-composition and taphonomy	74
2.2.5.2 Genetic species identification and DNA preservation through time	75
2.2.5.3 Within-site variation in DNA preservation	77
2.2.5.4 Between-site variation in DNA preservation	78
2.2.6 Concluding remarks	79
2.2.7 References	80
Chapter 3: Identifying microsatellites in the moa genome	83
Introducing the manuscripts in Chapter 3	84
Identification of microsatellites from an extinct moa species using high throughput (454) sequence data	86
3.1.1 Abstract	86
3.1.2 Introduction	86
3.1.3 Materials and methods	88
3.1.3.1 Sampling of moa fossils and DNA isolation	88
3.1.3.2 GS-FLX sequencing, microsatellite primer development and PCR	89
3.1.4 Results and discussion	90
3.1.5 Acknowledgements:	93
3.1.6 References	94
Profiling the dead: a characterization of six ancient DNA microsatellite markers for the extinct New Zealand moa – protocols, problems, and prospects	96
3.2.1 Abstract	96
3.2.2 Introduction	96
3.2.3 Materials and methods	98
3.2.3.1 Sampling and DNA extraction	98
3.2.3.2 High throughput sequencing	98
3.2.3.3 Microsatellite primer design and optimisation	99
3.2.3.4 Summary statistics, HW-tests and allelic dropout	100
3.2.3.5 Quantitative PCR	100
3.2.4 Results	101
3.2.4.1 High-throughput (GS-FLX) sequencing	101
3.2.4.2 Primer design and initial testing	101
3.2.4.3 Allelic dropout	104
3.2.4.4 Establishing the criteria	104
3.2.4.5 Descriptive statistics, HW-proportions and linkage	105
3.2.4.6 Quantitative PCR and dropout	105

3.2.5 Discussion	107
3.2.5.1 High-throughput sequencing	107
3.2.5.2 Allelic dropout	108
3.2.5.3 The data	110
3.2.6 Conclusion	111
3.2.7 References	112
Chapter 4: Mitochondrial DNA analyses of four sympatric moa populations	115
4.1 Introduction	116
4.2 Methods and results	118
4.2.1 ¹⁴ C chronology	118
4.2.2 Genetic diversity	121
4.2.3 Demography	125
4.2.4 An estimate of population size	130
4.2.5 The bigger picture – phylogeography	133
4.2.6 Assessing the robustness of the topology - a DensiTree analysis of <i>Dinornis robustus</i>	140
4.3 Discussion	142
4.3.1 Age	142
4.3.2 Genetic diversity	143
4.3.4 Demography and population size	144
4.3.5 Phylogeography	147
4.4 Conclusions	150
4.5 References	150
Chapter 5: A microsatellite study of the North Canterbury moa: population genetics in a temporal dimension	155
5.1 Introduction	156
5.2 Methods and results	160
5.2.1 Data generation	160
5.2.2 Tests for Hardy-Weinberg (HW) proportions and linkage disequilibrium (LD)	161
5.2.3 Genetic diversity	162
5.2.4 Heterozygosity through time	165
5.2.5 Genetic structure	165
5.2.6 Isolation by time	170
5.2.7 Genetic relatedness and kinship among <i>Dinornis robustus</i>	172
5.2.8 Identifying genetic outliers/migrants	174
5.2.9 Sex biased dispersal	177
5.3 Discussion	181
5.3.1 Basic descriptive statistics, tests for HW-proportions and LD	182
5.3.2 Genetic diversity through time	184
5.3.3 Genetic structure and isolation by time	185
5.3.4 Kinship analyses	186
5.3.5 Sex biased dispersal and genetic outliers	187
5.4 Concluding remarks	187
5.5 References	189
Chapter 6: General discussion	193
6.1 Future directions	199
6.2 Concluding remarks	202
6.3 References	202
Appendices	204

Acknowledgements

The past three years have been a long, intense, and incredibly exciting journey. My work has included lab-work in Australia, Denmark, and New Zealand, as well as excavations in New Zealand, and in remote caves in Australia. It has extended from the careful sampling and measuring of hundreds of ancient moa bones at Canterbury Museum and Te Papa, to many long nights, sweating over failing microsatellite PCR reactions in the aDNA lab in Perth. Even the final stages of writing proved far more “exciting” than anticipated, when a major earthquake hit Christchurch, just as I was putting together the final chapter.

First of all, I would like to thank the Marsden Fund of the Royal Society of New Zealand for funding the moa programme (Contract 06-PAL-001-EEB, Palaecol Research Ltd), including my scholarship.

Thanks to my three supervisors, Prof. Richard Holdaway, Dr. Michael Bunce and Dr. Marie Hale for being very supportive on both professional and personal levels. Thanks to Richard for giving me the possibility of a life-time by taking me on board, and thanks to Mike for all the insights on molecular biology and for having provided a great working environment at the aDNA lab (Murdoch University, Perth).

I am deeply grateful to Malene Møhl for her support throughout this adventure, and for putting >200 hours of her own time into this project. Her help and enthusiasm has been crucial to the entire project.

Thanks to Eske Willerslev, Tom Gilbert, Kenneth Andersen, Morten Rasmussen, Martin Hebsgaard, Astrid Schmidt Grene, and Pernille Selmer Olsen for a fruitful and ongoing collaboration and for letting me work in their laboratories in Copenhagen.

I thank the following friends and colleagues for having contributed: Paul Scofield, Alan Tennyson, Nicolas Rawlence, Charlotte Oskam, Carlo Pacioni, Nicole White, Jayne Houston, Helen Hunt, Emma McLay, Pieter Pelser, James Haile, Ross Barnett, John O’Brien, Frances Brigg, Alexei Drummond, Simon Ho, Jessie Wu, Gavin Prideaux, Remco Bouckaert, Peter Spencer, Christian Anderson, Deborah Dawson, Alison Johnston, Courtney Briggs, Steve Earle, and the Hodgen family.

Thanks to Kerry O’Connell and Paul South for taking good care of me, while being back in Christchurch for this final stretch. I will not forget that.

Finally, I owe big thanks to the Allentoft and Møhl families for their support and encouragement over the years.

Abstract

Ancient DNA (aDNA) was isolated from the bones of 290 individuals and four species of extinct New Zealand moa. All sampled bones had been recovered from a small geographic area (~10 km radius) near Waikari in North Canterbury. A total of 217 specimens were ^{14}C -AMS dated, providing a temporal framework for the genetic analyses and an unprecedented opportunity to study extinct megafauna at the population level. Taxon and sex were determined for each individual, using aDNA technology. This revealed a large excess of females (overall $\text{♂}:\text{♀} = 1:5.1$), and significant compositional differences for the moa assemblages between fossil sites. Balanced sex ratios were observed among juvenile moa, suggesting that a gender-bias developed as the birds matured, probably as a result of higher male mortality. Female territoriality and ecological niche-separation are discussed in this context. Mitochondrial DNA (mtDNA), amplified using a quantitative PCR procedure, provided a measure of DNA preservation in each radiocarbon-dated fossil. This assessment showed that DNA degrades over thousands of years according to an exponential decay model, and the average molecular half-life for the here targeted DNA fragment was estimated to be 521 years. By using high-throughput sequencing, six polymorphic moa microsatellite markers were identified and characterised. These are the first microsatellite primers developed exclusively for extinct taxa. A high-resolution genetic study of the four sympatric moa populations was carried out, combining information from mtDNA, microsatellites, sex-identification, and radiocarbon age. Genetic diversity, past demography, kinship, and other aspects of moa biology were analysed. The populations showed a remarkable extent of genetic stability throughout the 3000-4000 years preceding their extinction, suggesting that they were large and viable before suddenly disappearing. The results represent significant advances in aDNA research and thanks to the high resolution in microsatellite markers, moa have here been studied, almost as if they were still alive.

Chapter 1: Introduction

Eighteen years of ancient DNA research on extinct New Zealand moa (Aves: Dinornithiformes)

Following a brief introduction, this chapter will summarise molecular research on moa under three main categories (Evolution; Taxonomy; Biology) to provide background and context for the succeeding chapters. This is not intended as a detailed assessment of all relevant information but rather a concise review on moa genetics, as compiled in July 2010. I will finish this mini-review by discussing the potential for future research in this area.

1.1.1 Introduction

When Higuchi et al. (1984) isolated DNA from an extinct species of zebra (the quagga - *Equus quagga*) and Svante Pääbo shortly after repeated the achievement on tissue from Egyptian mummies (Paabo, 1985a; Paabo, 1985b), the field of ancient DNA (aDNA) was founded. Despite a set of serious challenges resulting from the low quantity and its degraded state (see Chapter 2, Section 2), the study of aDNA has proven a viable and productive science, continuing to yield publications in high impact journals and constantly pushing the boundaries in molecular biology. The information stored in the ancient bio-molecules can provide detailed insights on past biodiversity, prehistoric dispersal and colonization events, consequences of environmental changes, and extinction processes (Hofreiter et al., 2001; Willerslev, Cooper, 2005). Although there have been many highlights over the past 25 years, it has perhaps been DNA from the remains of large charismatic extinct animals (megafauna¹) that more than anything has defined the field, and attracted the attention of molecular biologists interested in the past. Ancient DNA has been extracted and analysed from extinct megafauna species such as woolly mammoth (*Mammuthus primigenius*) (e.g., Debruyne et al., 2008; Gilbert et al., 2008; Greenwood et al., 1999; Poinar et al., 2006), woolly rhino (*Coelodonta antiquitatis*) (e.g., Orlando et al., 2003; Willerslev et al., 2009), steppe bison (*Bison priscus*) (e.g., Shapiro et al., 2004), giant sloths (Megatheriidae) (e.g., Greenwood et al., 1999; Poinar et al., 1998) and cave bear (*Ursus spelaeus*) (e.g., Hofreiter et al., 2004; Orlando et al., 2002; Stiller et al., 2010) amongst others.

No other extinct taxon, however, has been studied so intensively with molecular technology as the New Zealand moa (Aves: Dinornithiformes) – at least in terms of number of publications and number

¹ “Megafauna” is a collective term used to describe a group of animals that weigh over 44 kg (Johnson, 2002) or alternatively; the component of the fauna of a region or period that comprises the larger terrestrial species, regardless of weight (Hansen and Galetti, 2009)

of profiled fossils. These wingless giant graviportal birds have continued to impress scientists as well as the public ever since the famous British anatomist Richard Owen was given the shaft of moa femur in 1839 (Figure 1.1.1) and predicted the existence of a large flightless bird in New Zealand (Owen, 1840).



Figure 1.1.1: Richard Owen standing next to a composite skeleton of a large female of the South Island giant moa (*Dinornis robustus*). In his right hand he is holding the bone fragment, which he based his hypothesis of the existence of moa on.

Image reference: Richard Owen, *Memoirs on the extinct wingless birds of New Zealand*. Vol. 2. London: John van Voorst, 1879, plate XCVII.

Like many native New Zealand birds, the herbivorous moa exhibited a *K*-selected life history, with delayed sexual maturity (Turvey et al., 2005; Turvey, Holdaway, 2005) and were therefore extremely vulnerable to the hunting and habitat changes that followed the arrivals of Polynesians in the late 13th century AD. (e.g., Anderson, 1989; McWethy et al., 2009). Moa hunter sites littered with bones and eggshell fragments can be found across New Zealand (Anderson, 1989) and represent the most compelling evidence that moa was hunted to extinction. The model presented in Holdaway and Jacomb (2000b) suggested that moa were driven to extinction within as little as 60 years after human colonisation, although it was later discussed whether isolated populations could have survived after this date (Anderson, 2000; Holdaway, Jacomb, 2000a).

After a century and a half of scientific investigations (see Worthy, Holdaway, 2002), moa research has now entered the era of molecular biology. The first study on ancient moa DNA was published in 1992 (Cooper et al., 1992) and since then, a considerable literature has added to our knowledge on these birds. Table 1.1.1 lists all the aDNA studies on moa published to date. These references represent a wide array of topics and a range of molecular techniques applied to different biological samples of moa origin, including mummified tissue (Cooper et al., 1992), bone (e.g., Cooper et al.,

2001), coprolites (Wood et al., 2008a), feathers (Rawlence et al., 2009), eggshell (Oskam et al., 2010), and ancient sediments containing moa DNA (Haile et al., 2007; Willerslev et al., 2003). From brief critical comments on moa nomenclature derived from genetic analyses (Worthy, 2007b) to large-scale sequencing projects involving whole mitochondrial genomes (Cooper et al., 2001; Haddrath, Baker, 2001) or DNA sequences from hundreds of individuals (Baker et al., 2005; Bunce et al., 2009). Not all the research reported in Table 1.1.1 has generated or directly analysed DNA sequences, but they do at least relate to, or somehow rely on, genetic moa data in reaching their conclusions. Everything published on moa since 1992 has been drafted with the genetic information available, so it is not easy to draw a definite line between DNA-based and non DNA-based moa research. Among the contributions in this grey area (not listed in Table 1.1.1), are two books on the extinct New Zealand avifauna (Tennyson, Martinson, 2006; Worthy, Holdaway, 2002) and two recent comprehensive review papers on molecular phylogeographic patterns in New Zealand (Trewick, Gibb, 2010; Wallis, Trewick, 2009).

The 27 papers listed in Table 1.1.1, cover 18 years of science. With five publications appearing in 2009/10, it is clear that moa genetics is still a vigorous area of research. This may seem like a considerable effort expended on investigating DNA from one particular extinct lineage of birds, and it certainly reflects the fascination that surrounds the moa. However, it also reflects the relative abundance of well-preserved moa material. DNA degrades with time (see Chapter 2, Section 2) so the recent extinction event, coupled with a relatively cool climate (temperate) in New Zealand, means that moa genetics is perhaps slightly less challenging than experienced with other ancient taxa from non frozen environments. Also, on a planet currently experiencing anthropogenically mediated mass extinctions it is pertinent to gain insights on extinction processes. The moa undoubtedly represent the best known possibility to study the extinctions of large vertebrates at the molecular level and adding new detailed information to the extinction debate has been, and still is, an important incentive to study moa (see also ‘Aims and Scopes’ section, category 4).

1.1.2 Moa evolution and phylogeography

Modern birds are separated taxonomically into palaeognaths and neognaths based on characters of the bony palate which is considered more archaic in the former group. This basal split is strongly supported by genetic analyses (Hackett et al., 2008; Paton et al., 2002; Slack et al., 2007). Paleognaths include the volant tinamous (Tinamidae) from South America and the flightless ratites, traditionally considered to form sister groups. The modern ratites include the ostrich (*Struthio camelus*) from Africa, the Australian emu (*Dromaius novaehollandiae*) and cassowaries (*Casuarius* spp.), the South American rheas (*Rhea* spp.), the kiwi (*Apteryx* spp.) from New Zealand and the recently extinct moa

(Dinornithiformes) and elephant birds (*Aepyornis* spp., *Mullerornis* spp), from New Zealand and Madagascar respectively. All ratites have highly reduced wings and in moa these have been completely lost. DNA technology has been applied vigorously to study ratite evolution, providing new insights on the phylogeny, the phylogeography, and the timing and events causing radiation within the ratites and specifically within the moa lineage.

New Zealand has a very complex geological and biogeographical history with a series of major events that have shaped the environment and biota, starting with the gradual Cretaceous separation of proto New Zealand (Zealandia) from Gondwana about 82 million years ago (Mya). The Oligocene drowning reduced the low-lying landmass to a series of islands ~25 Mya, followed by the land rise and the formation of the Alpine Fault 5-7 Mya. Later, several major volcanic eruptions and >50 glacial cycles throughout the Pleistocene caused massive environmental disturbances (summarised in Gibbs, 2006). New Zealand was also the last major landmass to be inhabited by humans. For a biota adapted to complete absence of mammals (except for a few species of bats), human colonisation proved catastrophic, resulting in a long list of extinctions (McGlone et al., 2003; Tennyson, Martinson, 2006; Worthy, Holdaway, 2002), distribution rearrangements, and the introduction of invasive species (reviewed in Hamel et al., 2003). The potential for studying evolutionary patterns, phylogeography, and past population demographics of taxa in New Zealand is exceptional, but the patterns are at the same time extremely complex.

With the acceptance of plate tectonics theory in the 1960s, the vicariance model of allopatric speciation offered an elegant explanation for the strange New Zealand biota – it had been shaped by isolation after Zealandia drifted from Gondwana ~82 Mya. The distribution of ratites on the different continents of the Southern Hemisphere has often been cited as a prime example of vicariance, with flightless ancestors being isolated on the landmasses as they slowly drifted apart (e.g., Cracraft, 1974). However, recent research, including molecular data, have demonstrated that this model is too simplistic, by suggesting dispersals and radiations postdating the break-up of Zealandia-Gondwana (reviewed in Trewick, Gibb, 2010; Wallis, Trewick, 2009). Studies on moa have played a central role in this discussion.

Indeed, the first aDNA study on moa (Cooper et al., 1992) showed that a simple vicariance model was not applicable. By extracting DNA from mummified moa specimens and comparing their mitochondrial DNA (mtDNA) 12S rRNA genes with those in extant ratites, it was demonstrated that moa and kiwi are not monophyletic. A common moa-kiwi ancestor could therefore not have been isolated on Zealandia by continental drift. Rather, two independent colonisation events had taken place or ratites radiated even before Zealandia drifted from Gondwana (or both). Later, two studies based on whole mitochondrial genomes (Cooper et al., 2001; Haddrath, Baker, 2001) confirmed that moa were more basal in the ratite phylogeny than kiwi, and suggested that a kiwi ancestor would have

crossed the Tasman Sea to arrive in New Zealand later than the moa. Both these studies also attempted to date the speciation events, using a molecular clock approach. Cooper et al. (2001) used the 82 Mya Zealandia-Gondwana split as a calibration point for separating the moa lineage, which can seem logically flawed. If the Tasman Sea was not an efficient barrier for the kiwi ancestor, there is no particular reason to believe that it would have been a barrier for the moa ancestor millions of years before. Haddrath and Baker (2001), on the other hand, calibrated their molecular clock using an estimated 35 million-year-split between emu and cassowary, based on a single Australian fossil. Despite the different approaches, the two studies reached similar conclusions. Haddrath and Baker (2001) actually provided an independent confirmation of the otherwise problematic moa split around 80 Mya. Both groups placed the kiwi speciation to somewhat later, at ca. 60-70 Mya. Not only a late kiwi separation point was problematic from a vicariance perspective, but also the ostrich appeared to split off long after the African continent drifted from Gondwana. Despite these exceptions from the model, the two groups still discussed their results in a vicariance context.

More recent molecular contributions have shed further light on the problem and the concept of ratite radiation by vicariance has now largely been rejected. In two comprehensive phylogenomic analyses of multiple nuclear loci from numerous bird species, Hackett et al. (2008) and Harshman et al. (2008) showed that ratites are not monophyletic. The volant South American tinamous, previously considered as sister group to the ratites, are nested within the ratite clade. A topology showing volant bird species within lineages of flightless ones requires either that the loss of flight happened several times or that flight was lost in an archaic palaeognath and then regained among the tinamous. The authors advocate the former scenario, recording multiple instances of loss of flight in other bird lineages but none of flight being regained (Harshman et al., 2008).

If true, this means that the traditional ratite clade is paraphyletic and that the ratite monophyly enforced (by excluding tinamous) in previous analyses (Cooper et al., 2001; Paton et al., 2002; Pereira, Baker, 2006) is incorrect. Phillips et al. (2010) has the final word so far, by providing a solid confirmation of the ratites as paraphyletic. Whereas Hackett et al. (2008) and Harshman et al. (2008) used nuclear loci of extant birds, Phillips et al. (2010) repeated the two whole mt-genome analyses on moa, but included a series of additional avian mtDNA genomes and crocodiles. Using complex phylogenetic analyses, involving model partitioning between different codons and genes, and accounting for base composition bias, they showed that moa and tinamou are sister groups within the ratites. The analyses estimated that moa and tinamou separated about 53 Mya (95% HPD = 39.1-71.5 Mya), and it was suggested that a moa ancestor flew to New Zealand. Flight appears to have been lost 4 or 5 times within the ratites, eliminating the need to invoke vicariance in models of ratite distribution and radiation. Phillips et al. (2010) also suggest that Gondwana might not even represent the place of palaeognath origin.

In addition to the studies focusing on ratite evolution and placement of the moa lineage, a number of papers on aDNA have specifically investigated speciation and phylogeography *within* the moa lineage. The main contributions so far are Baker et al. (2005) and Bunce et al. (2009) who both used mtDNA sequences from hundreds of individuals in attempts to define species boundaries and to reconstruct the timing and catalysts for moa radiation. Baker et al. (2005) identified 14 distinct genetic lineages within moa, and their analyses suggested that moa radiation started with the separation of *Megalapteryx* about 18.5 Mya, followed by *Dinornis* about 15 Mya (see next section on taxonomy). After that, several lineage-splits occurred in the 4-10 Mya interval, likely related to the formation of the Southern Alps and the separation of the two main islands. The analyses and interpretations in Baker et al. (2005) may be compromised by the use of the 82 Mya Gondwana-Zealandia split as a calibration point for the molecular clock. As outlined above, this calibration point seems to be at least conceptually invalid because early ratite may have been volant. Bunce et al. (2009) took another approach, using two independent molecular clock analyses. One analysis included external calibration points based on well accepted avian fossil data (excluding the 82 Mya point) and one included internal calibration based on radiocarbon dates of the bones associated with the aDNA sequences. In their analyses, *Megalapteryx* still emerged as the basal moa lineage, but the split was now estimated to have happened just 5.8 Mya, followed by an extensive radiation associated with the mountain building in this period (Figure 1.1.2). Another interesting result was that the genus *Euryapteryx* did not display North-South Island monophyly, which suggested that whatever barrier was separating the two islands, it did not prevent gene flow for this particular species (Figure 1.1.3).

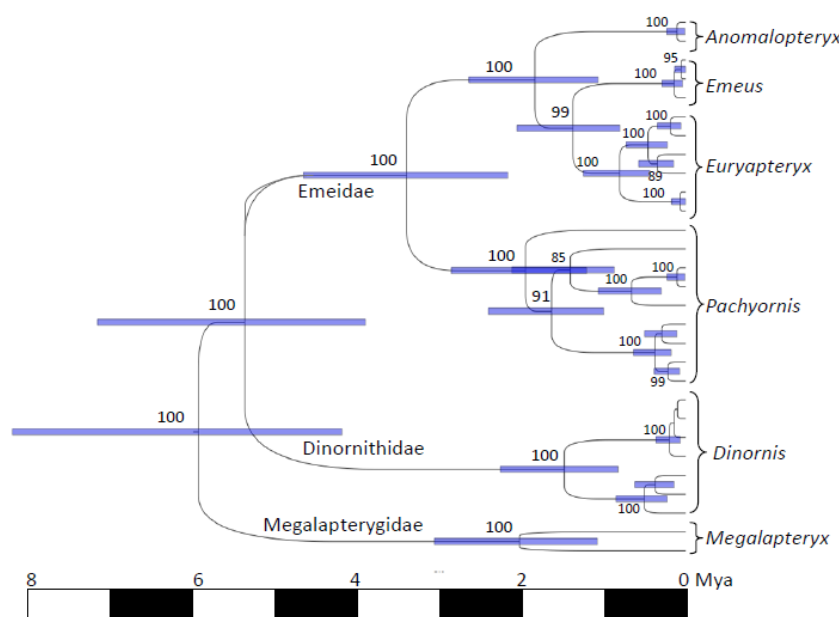


Figure 1.1.2: Phylogenetic reconstruction of the six moa genera (based on 2153 bp of coding mtDNA from 29 moa), including a proposed timescale for the radiation. Numbers on the nodes represent posterior probabilities as estimated in the Bayesian analyses applied by the authors. Bars represent 95% probability distribution for the timing of each split.

Image reference: Modified from Bunce et al. (2009).

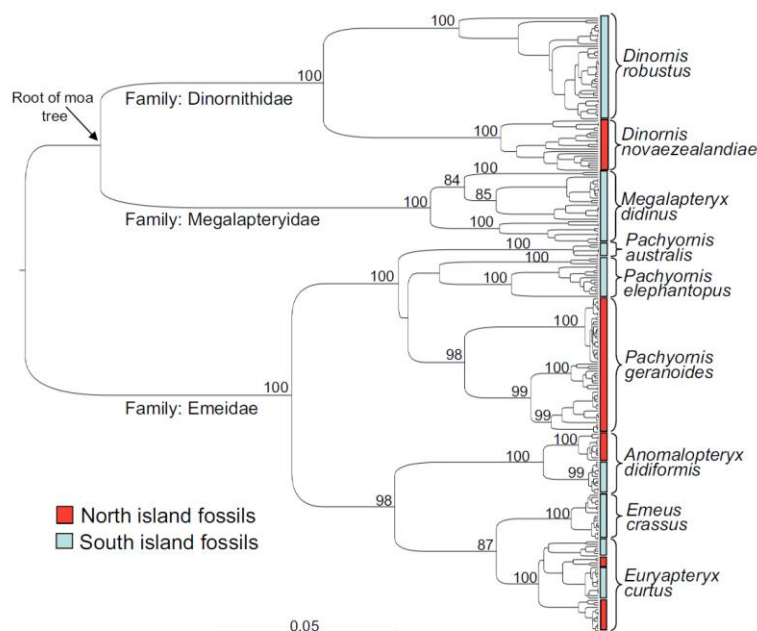


Figure 1.1.3: Phylogenetic reconstruction of the nine presently recognised moa species (based on 389 bp of control region mtDNA from 263 moa). Note that the genus *Euryapteryx* does not display North-South island monophyly. The root is here placed incorrectly because of saturation effects in this data.

Image reference: Bunce et al. (2009).

This chapter focuses explicitly on the molecular aspects of moa research but two recent non-molecular studies are worth mentioning because they display the controversy still prevalent in the study of ratite and moa evolution. Bourdon et al. (2009) conducted a comprehensive phylogenetic study based on 17 species and 129 morphological characters and suggested that moa and kiwi are monophyletic, in strong opposition to all the molecular data. A potential major source of error was their use of tinamous as outgroup in the cladistic analyses. If this is incorrect, as is suggested by the DNA, derived and primitive characters could have been confused.

Regarding the timing of moa radiation; Tennyson et al. (2010) described the oldest moa fossils (eggshell and bone fragments) recovered so far, extending the presence of a “proto-moa” in New Zealand back to the Miocene, at least 16-19 Mya. Moreover, differences in the eggshell thickness indicate the presence of several moa taxa before the formation of the Alpine Fault. This does not, however, contradict ~5 Mya radiation suggested in Bunce et al. (2009). The authors in the latter study specifically state that although the Quaternary moa fauna we are familiar with seem to stem from a single lineage, it could just be the latest of several moa radiations, with some lineages subsequently going extinct, for example as a consequence of the Oligocene drowning.

1.1.3 Taxonomy and genetic identification

Genetic analyses have greatly assisted in the clarification of species boundaries within moa. Although there are many unresolved questions, DNA technology has provided a solid foundation for investigations of moa diversity. The present consensus is nine species (Bunce et al., 2009; Tennyson, Martinson, 2006) (Figure 1.1.4), ranging from the smallest 12-15 kg turkey-sized, North Island form of *Euryapteryx curtus* to the female of *Dinornis robustus* weighing >200 kg and being two meters over the back. However, 64 species and 20 genera names have been proposed since Owen announced the presence of a struthious bird in New Zealand (Worthy, Holdaway, 2002), reflecting the difficulties in establishing diagnostic morphological characters for accurate species determination. There are several reasons for these problems. First, the New Zealand fossil record includes nine superficially similar ratites, many of them sympatric, so it was challenging for early palaeontologists to discriminate between the different species without large comparative collections. Second, moa display significant sexual size dimorphism, and last; intra-specific size variation among fossils from different geographic regions (Worthy, 1987; Worthy, Holdaway, 1996) undoubtedly contributed to the confusion. Together, these factors are responsible for a whole continuum of adult body size, constituting a true taxonomic minefield.

Although much of this confusion had already been untangled before the advent of DNA (Cracraft, 1976; Worthy, Holdaway, 2002), the studies of Bunce et al. (2003) and Huynen et al. (2003) represent very illustrative examples on the potential dangers in relying solely on morphological characters to establish taxonomic boundaries. By targeting female specific W-chromosomal DNA fragments, it was shown that presumed species within *Dinornithidae*, largely separated on the basis of size, were in fact male and female *Dinornis* (Figure 1.1.5). This work necessitated a revision of moa taxonomy and nomenclature.

The molecular relationships of moa were further refined in the two large phylogenetic studies by Baker et al. (2005) and Bunce et al. (2009), but the number of species has remained controversial. Baker (2007a) proposed species status for all 14 moa lineages, arguing that the genetic distances in the COI-gene (presented in Lambert et al., 2005), could be used to apply a bar-coding approach on moa and other extinct taxa. However, the impact of this proposal was diluted by problems with the applied nomenclature (Baker, 2007b; Worthy, 2007a). Bunce et al. (2009) took a more conservative approach towards recognising species boundaries and did not rely solely on the signal in the genetic data. They recognised three moa families (*Dinornithidae*, *Emeidae*, *Megalapterygidae*), six genera, and nine species (Figures 1.1.2; 1.1.3 and 1.1.4) and the authors rejected the use of a strict bar-coding approach as advocated elsewhere (Baker, 2007a; Lambert et al., 2005).

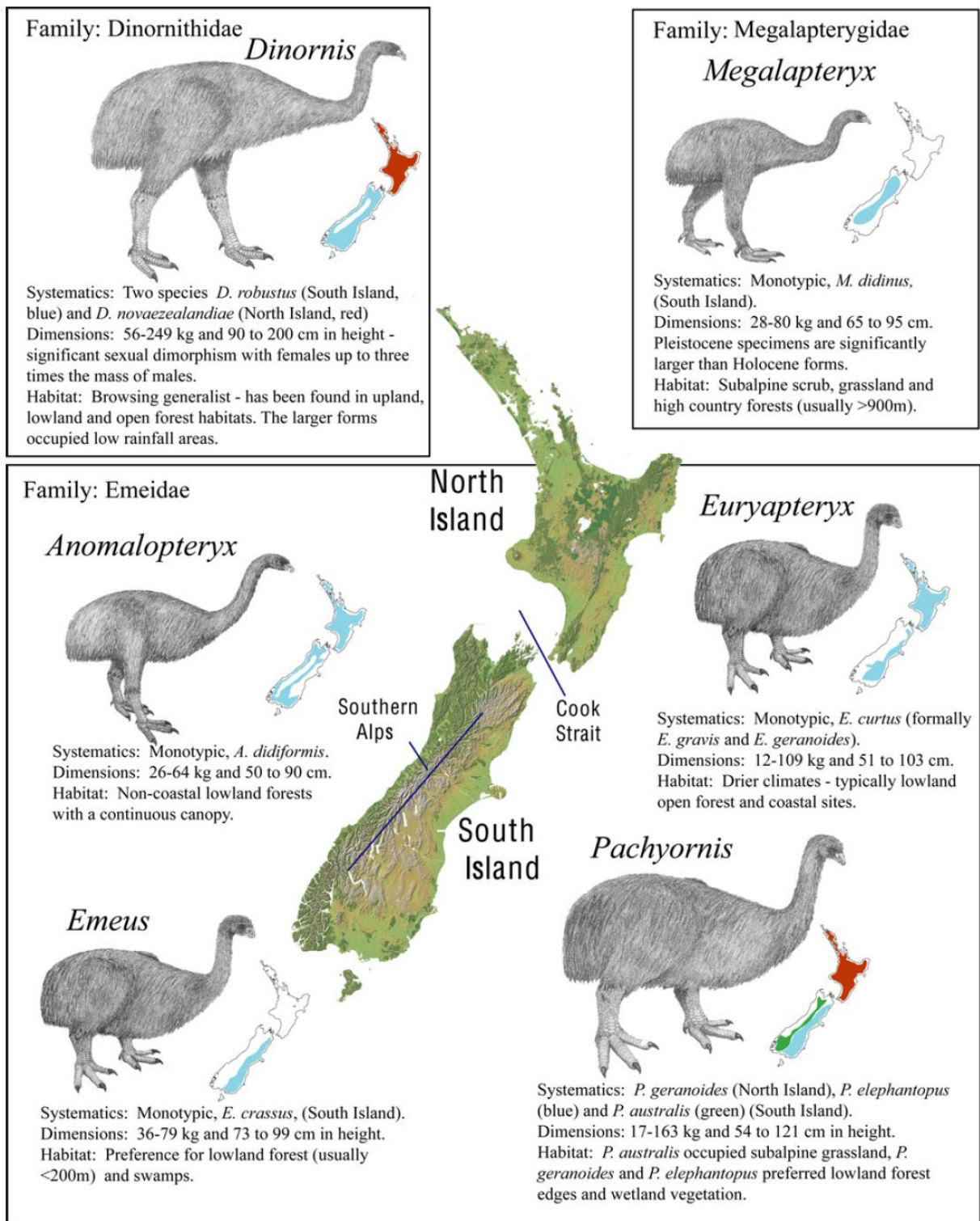


Figure 1.1.4: A summary of systematics, dimensions, and estimated distributions of the three families, six genera, and nine species of moa, as proposed in Bunce et al. (2009).

Image reference: Bunce et al. (2009).

Their arguments against bar-coding focused on the difficulties in distinguishing between temporal, geographic and reproductive genetic barriers when investigating DNA from closely-related extinct animals. Also, if female moa held a home range (see Chapter 2, Section 1), the mtDNA data might represent a skewed image of the intra- contra interspecific genetic diversity. As more information becomes available from DNA sequences, radiocarbon dates and isotopic data, a better understanding of moa ecology and the temporal and spatial patterns that shaped moa gene pools will emerge. It is unlikely that moa phylogeny and taxonomy will stabilise in the near future. Several areas of moa research have benefitted considerably from the genetic data generated since 1992. More than 300 moa mtDNA Control region sequences of various lengths are now available from GenBank, and their value has been demonstrated in a series of recent publications, where the DNA was extracted to obtain a secure species ID rather than providing phylogenetic information *per se*. This is relevant when studying biological substrata such as moa coprolites (Wood et al., 2008a), feathers (Rawlence et al., 2009), or eggshell (Oskam et al., 2010) which might lack obvious taxon-informative characters (see below). Moreover, identifying moa bones can still represent considerable challenges even to trained palaeontologists. For example, bone fragments or bones from juveniles usually lack diagnostic features of the adult bone, and the (smaller and rarer) males are easily misidentified as a smaller taxon. As chapter 2 (Section 1) in this thesis demonstrates, despite low curatorial error rates (Scofield et al., 2005), any taxonomic assessment of a moa fossil assemblage will benefit from a genetic confirmation of results, if the DNA preservation permits. Particularly if juvenile bones and/or fragmented bones are present.



Figure 1.1.5: Extreme reverse sexual size dimorphism in *Dinornis robustus*. The females were up to 150% taller and 280% heavier than the males. A pigeon skeleton is shown for size comparison.

Image reference: Allentoft et al. (2010).

1.1.4 Moa biology

By using ancient DNA for identification rather than analysing genetic diversity *per se*, several publications have shifted the focus from larger evolutionary questions towards examining moa biology and the environment they occupied. Wood et al. (2008a) extracted DNA from 24 moa coprolites recovered from rock shelters in Otago, South Island. Mitochondrial DNA was amplified to determine the moa species associated with each coprolite, and the coprolite contents were then examined microscopically to identify the plant remains. The analyses showed that coprolites of four species were present in the sample, and that moa sampled a diverse range of plants (at least 30 taxa) dominated by herbs and shrubs <30 cm tall. The study did not reveal any evidence of dietary niche partitioning, showing almost complete overlap in plant preferences between species, despite their obvious differences in beak-shapes and body sizes. This result is against the expectations and warrants further studies on a larger sample set.

Ancient DNA has also been used to identify moa feathers. Rawlence et al. (2009) extracted DNA from both the shafts and the tips of ancient feathers, which allowed an investigation of plumage dimorphism in moa. Their study showed that three of four moa species in the area had brown feathers with black tips, whereas feathers from *Pachyornis elephantopus* were shorter and had white tips, creating a speckled plumage pattern. These colour deviations could perhaps reflect slightly different habitats, prompting different camouflage against predation from the massive and now extinct Haast's Eagle (*Hieraaetus moorei*).

In the only attempt so far to estimate moa population sizes from genetic information, Gemmell et al. (2004) used all the *Dinornis* mtDNA Control region sequences available on GenBank at that time. The level of genetic diversity was estimated, and based on a theoretical correlation with effective population size, the latter was calculated. The standing population of *Dinornis*, in interval 6000 BP to 1000 BP, was estimated to be 930,000 (95% CI = 720,000-1,668,000) in the North Island and 479,000 (95% CI = 321,000-694,000) in the South Island. These numbers are an order of magnitude larger than the 159,000 estimated for all species at both Islands combined, at the time of human arrival, based on analyses of fossil distribution, abundance and radiocarbon ages (Holdaway, Jacomb, 2000b). Rather than seeing a contradiction, Gemmell et al. (2004) concluded that *Dinornis* must have experienced a massive decline in population size just before human settlement to reach the low numbers estimated by Holdaway and Jacomb (2000b). While an interesting proposition, the genetic analyses were based on a simplistic approach, using data that probably violated the assumptions of the methodology. For example, it was assumed that the GenBank sequences represented a panmictic *Dinornis* population, although the fossils, from which the DNA was extracted, covered thousands of years and large geographic distances. The approach is revisited and discussed in more detail in Chapter 4.

1.1.5 The future of moa genetics

Last, I will assess potential future directions for molecular research on moa. This topic will be revisited in the general discussion (Chapter 6), relating more specifically to the research in this thesis.

Although the moa gene pool is one of the most extensively investigated of any extinct taxon, this does not imply that all possible information has been extracted. The 18 years of aDNA research has generated as many new questions as answers, and long standing problems remain, such as the early evolutionary splits in the birds and the radiation of ratites. New high-throughput sequencing (HTS) platforms (see Chapter 3) are generating enormous amounts of data, and ever more sophisticated methods for handling and analysing large genetic datasets may revise or refine the current knowledge. As outlined above, the most recent molecular publications on this matter appear to be approaching a consensus of ratites as paraphyletic and vicariance as insufficient for explaining the radiation. However, this viewpoint is not shared unequivocally among morphologists, and more studies that attempt to integrate information from both fields would be beneficial. This could involve analyses of genetic variation in known functional genes combined with assessments of the variation in associated morphological traits.

High throughput sequencing might also have a significant role to play in clarifying moa evolution, for example by allowing the sequencing of many entire mt-genomes from all the identified moa lineages. Such data should provide a strong basis for a re-evaluation of the taxonomy. A recent phylogeographic study of killer whales (*Orcinus* spp.) showed the potential of the technique. By sequencing entire mt-genomes of 139 individuals (Morin et al., 2010), a highly informative dataset was generated. Although an equivalent study on fragmented moa DNA would be a much bigger challenge, there is little doubt that it can be done with the available technology, as demonstrated by the 18 complete mammoth mt-genomes analysed in Gilbert et al. (2008). In time, the whole nuclear moa genome could be sequenced, given the excellent preservation and the recent advances in ancient genomics – exemplified with the publication of the entire genome of a 4000 year old Greenlandic Eskimo (Rasmussen et al., 2010) and the Neanderthal genome (Green et al., 2010). However, before a ratite genome becomes available as a scaffold, the problems could easily overshadow the benefits of attempting a full *de novo* assembly of moa sequences.

As the papers listed in Table 1.1.1 clearly shows, it does not require whole genomes and massive sequencing efforts to extract new and valuable information from extinct animals. Simple genetic identifications of feathers, coprolites, eggshells, and bones have led to new insights on moa biology and will do so in the future. Eggshell, in particular, are often found in archaeological contexts in the ovens and middens of the earliest Polynesian settlements (Gill, 2010; Keepax, 1981), and apparently preserves aDNA better than bone. This biological substrate represents a valuable and largely untapped molecular resource for studying the human-moa interactions and the extinction process (Oskam et al.,

2010). Although a broad outline of moa phylogenetics is now available, it is still unclear how the genetic signals within each moa species should be interpreted. For example, were the deep lineages in the South Island *Dinornis* (Figure 1.1.3) a result of gene flow barriers in space, or time, or both? Additional DNA sequences with associated radiocarbon dates are necessary to resolve such issues. The molecular work on moa, presently underway in several aDNA laboratories, is slowly filling in the gaps in our temporal and spatial knowledge, by contributing new DNA sequences each year. For example, the sequences associated with this thesis (Digital Appendix 1) will, when submitted to GenBank, effectively double the number of moa control region sequences available.

The last of the potential advances dealt with here, is moa population genetics – the core topic of this thesis. Many thousands of moa bones have been recovered since the first descriptions in 1839. In particular, caves and swamps have yielded late Pleistocene- and Holocene-aged fossils in very high concentration (e.g., Wood et al., 2008b; Worthy, Holdaway, 1996; Worthy, Holdaway, 2002). These settings offer the very rare possibility in aDNA research of studying extinct megafauna at the population level. By applying HTS, fully functional and highly variable microsatellites markers have been characterised for moa (see Chapter 3), and it should now be possible to generate a detailed image of the genetic diversity that characterised the moa populations during the late Holocene: Life before the end.

Table 1.1.1: Chronological overview of 27 publications on moa aDNA

Reference	Title	Substrate	Target DNA	Main findings
Cooper et al. (1992)	Independent origins of New Zealand moas and kiwis	skin, muscle, bone	mtDNA (12S)	The first moa aDNA paper. Demonstrates that moa and kiwi are not monophyletic, so two independent ratite colonizations of New Zealand must have taken place.
Cooper (1993)	DNA from museum specimens	as above + feather	mtDNA (12S)	Paper discusses in detail the methods applied in Cooper et al. (1992) to successfully isolate DNA from ancient moa remains.
Cooper & Cooper (1995)	The Oligocene bottleneck and New Zealand biota: genetic record of a past environmental crisis	as in Cooper (1993)	mtDNA (12S, ND6)	Moa contain less genetic variability than other bird lineages. Same pattern is apparent in Kiwi , and NZ wrens, suggesting a massive prehistoric bottleneck - likely facilitated by the Oligocene drowning.
Vickers-Rich et al. (1995)	Morphology, myology, collagen and DNA of a mummified upland moa, <i>Megalapteryx didinus</i> (Aves: Dinornithiformes) from New Zealand	skin muscle	moa DNA, no target	Discuss DNA extraction from a mummified moa and show in hybridisation experiments that the extracted DNA is not contamination.
Cooper and Penny (1997)	Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence	GenBank	mtDNA (12S)	DNA from many birds (including moa), was analysed to demonstrate that at least 22 avian lineages survived the Cretaceous-Tertiary boundary, which was the end for many other vertebrates.
Cooper (1997)	Studies of Avian ancient DNA: From Jurassic Park to modern island extinctions	as in Cooper et al. (1992)	mtDNA (12S + ND6)	A re-analyses of the data presented in Cooper et al. (1992). More data are included and confirm that moa and kiwi are not monophyletic.
Cooper et al. (2001)	Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution	bone	mtDNA (whole genomes)	First mt-genomes of extinct taxa. Used to clarify ratite evolution. The analyses suggest that rhea is basal and that the moa lineage splits off next, as sistergroup to the other ratites.
Haddrath & Baker (2001)	Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis	bone	mtDNA (whole genomes)	A ratite evolution study based on whole mt-genomes. The analyses suggest that moa is basal and that the rhea splits off next as sistergroup to the other ratites.
Paton et al. (2002)	Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds	GenBank	mtDNA (whole genomes)	Two moa mt-genomes are included in a big dataset, to reject a hypothesis of shorebirds as ancestors to modern birds. Results show palaeognaths as sistergroup to all other modern birds (including shorebirds)
Bunce et al. (2003)	Extreme reversed sexual size dimorphism in the extinct New Zealand moa <i>Dinornis</i>	bone	nuclear sex markers + mtDNA (CR)	Moa females were much larger than males and previous morphological taxon designation in Dinornithidae simply reflects different genders.
Huynen et al. (2003)	Nuclear DNA sequences detect species limits in ancient moa	bone	nuclear sex markers + mtDNA (CR)	Shows that moa display extreme reverse size dimorphism, and this genetic insight is used to define new species limits.
Willerslev et al. (2003)	Diverse plant and animal genetic records from Holocene and Pleistocene sediments	sediment	mtDNA (CR + 12S)	First study documenting that aDNA can be extracted from ancient sediment. The study includes moa sequences amplified from New Zealand sediment samples.
Gammel et al. (2004)	Moa were many	GenBank	mtDNA (CR)	A genetic estimate of moa pop. sizes is compared to a previous estimate based on ecological data, and the difference is interpreted as evidence that moa were in decline prior to human arrival in New Zealand.

Baker et al. (2005)	Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: the giant moas of New Zealand	bone	mtDNA (CR + more)	Uses molecular clocks to date moa radiation. Detects 14 monophyletic lineages. First taxon to split off is <i>Megalapteryx</i> spp. 18.5 mya. Remaining moa radiation happened 4-10 mya.
Lambert et al. (2005)	Is a large-scale DNA-based inventory of ancient life possible?	bone	mtDNA (CR + 12S + COI)	Investigates whether the COI gene is suitable for barcoding in moa and the presented levels of intra- and interspecific genetic differentiation support the applicability.
Baker (2007a)	Molecular advances in the study of geographic variation and speciation in birds	GenBank	mtDNA (CR + 12S + COI)	Discusses geographic genetic variation and speciation for a range of New Zealand taxa. Advocates full species status for all 14 moa lineages identified in Baker et al. (2005).
Baker (2007b)	Nomenclatural errors in moa taxonomy: a reply to Worthy	NA	NA	A reply to the critique in Worthy (2007). See below
Haile et al. (2007)	Ancient DNA chronology within sediment deposits: are paleobiological reconstructions possible and is DNA leaching a factor?	sediment	mtDNA (CR)	Shows that moa DNA can be extracted from sediment. Study also suggests that some downwards DNA migration, obstructing molecular stratification, happens for sheep DNA but not for moa.
Worthy (2007)	Moas and phylogenomics: How nomenclatural errors do a disservice to the understanding of moa taxonomy	NA	NA	Criticises the nomenclature applied by Baker (2007) in his advocacy for full species status of all 14 genetic lineages of moa.
Huynen et al. (2008)	Genetic identification of moa remains recovered from Tiniroro, Gisborne	bone	mtDNA (CR)	Confirming that mtDNA can be extracted from moa bones and bone fragments
Wood et al. (2008)	Coprolite deposits reveal the diet and ecology of the extinct New Zealand megaherbivore moa (Aves, Dinornithiformes)	coprolites	mtDNA (CR)	aDNA was extracted to identify 24 moa coprolites. Plant content was analysed (non-DNA) showing a diverse diet of low shrubs. No obvious signs of dietary niche separation between species.
Allentoft et al. (2009)	Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data	bone	nuclear STRs	A "proof of concept" paper showing how microsatellite markers can be developed from aDNA.
Bunce et al. (2009)	The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography	bone	mtDNA (CR + more)	Establishes the current consensus of 9 moa species. Phylogeography and NZ geology are discussed. Moa radiation did not happen until 5.8 mya with <i>Megalapteryx</i> splitting from the rest.
Rawlence et al. (2009)	DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa	feather	mtDNA (CR)	Shows that DNA can be extracted from very old (moa) feathers. The level of color preservation is investigated, followed by attempts to reconstruct moa plumage.
Allentoft et al. (2010)	Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand's extinct megafauna	bone	nuclear sex markers + + mtDNA (CR) + nuclear STRs	A moa population study. Extreme excess of females among fossils from adults, but not among juveniles. Large compositional differences (taxa, sex, maturity) between sites despite proximity in space and time.
Oskam et al. (2010)	Fossil avian eggshell preserves ancient DNA	eggshell	mtDNA (CR)	Shows that aDNA can be extracted from fossil eggshell, including five moa species.
Phillips et al. (2010)	Tinamous and moa flock together: mitochondrial genome sequence analysis reveals independent losses of flight among ratites	GenBank	mtDNA (whole genomes)	By removing enforced monophyly of ratites, it is shown that ostrich is basal in the phylogeny and that (volant) tinamous and moa are sister groups. This implies that flight was lost several times.

1.1.6 References

- Allentoft ME, Bunce M, Scofield RP, Hale ML, Holdaway RN (2010) Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand's extinct megafauna. *Quaternary Science Reviews* **29**, 753-762.
- Allentoft ME, Schuster S, Holdaway RN, et al. (2009) Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *Biotechniques* **46**, 195-200.
- Anderson A (1989) *Prodigious Birds* Cambridge University Press, Cambridge, England.
- Anderson A (2000) Less is moa. *Science* **289**, 1472-1473.
- Baker AJ (2007a) Molecular advances in the study of geographic variation and speciation in birds. *Ornithological Monographs* **63**, 18-29.
- Baker AJ (2007b) Nomenclatural errors in moa taxonomy: A reply to Worthy. *Auk* **124**, 1449-1450.
- Baker AJ, Huynen LJ, Haddrath O, Millar CD, Lambert DM (2005) Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: The giant moas of New Zealand. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8257-8262.
- Bourdon E, de Ricqlès A, Cubo J (2009) A new Transantarctic relationship: morphological evidence for a Rheidae-Dromaiidae-Casuariidae clade (Aves, Palaeognathae, Ratitae). *Zoological Journal of the Linnean Society* **156**, 641-663.
- Bunce M, Worthy TH, Ford T, et al. (2003) Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425**, 172-175.
- Bunce M, Worthy TH, Phillips MJ, et al. (2009) The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20646-20651.
- Cooper A (1993) DNA from museum specimens. In: *Ancient DNA: recovery and analysis of genetic material from paleontological, archaeological, museum, medical, and forensic specimens* (eds. Herrmann B, Hummel S), pp. 149-165. Springer Verlag, New York.
- Cooper A (1997) Studies of avian ancient DNA: from Jurassic park to modern island extinctions. In: *Avian molecular evolution and systematics* (ed. Mindell DP), pp. 345-373. Academic Press.
- Cooper A, Cooper RA (1995) The Oligocene bottleneck and New Zealand biota - genetic record of a past environmental crisis. *Proceedings of the Royal Society of London Series B-Biological Sciences* **261**, 293-302.
- Cooper A, Lalueza-Fox C, Anderson S, et al. (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704-707.
- Cooper A, Mourer-Chauvire C, Chambers GK, et al. (1992) Independent origins of New Zealand moas and kiwis. *Proceedings of the National Academy of Sciences* **89**, 8741-8744.
- Cooper A, Penny D (1997) Mass survival of birds across the Cretaceous-Tertiary boundary: Molecular evidence. *Science* **275**, 1109-1113.
- Cracraft J (1974) Phylogeny and evolution of ratite birds. *Ibis* **116**, 494-521.
- Cracraft J (1976) The species of moa (Aves: Dinornithidae). *Smithsonian Contributions to Paleobiology* **27**, 189-205.
- Debruyne R, Chu G, King CE, et al. (2008) Out of America: Ancient DNA evidence for a New World origin of Late Quaternary woolly mammoths. *Current Biology* **18**, 1320-1326.
- Gemmell NJ, Schwartz MK, Robertson BC (2004) Moa were many. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, S430-S432.

- Gibbs G (2006) *Ghosts of Gondwana. The history of life in New Zealand* Craig Potton Publishing, Nelson, New Zealand.
- Gilbert MTP, Drautz DI, Lesk AM, et al. (2008) Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 8327-8332.
- Gill BJ (2010) Regional Comparisons of the thickness of moa eggshell fragments (Aves: Dinornithiformes). *Records of the Australian Museum* **62**, 115-122.
- Green RE, Krause J, Briggs AW, et al. (2010) A Draft Sequence of the Neandertal Genome. *Science* **328**, 710-722.
- Greenwood AD, Capelli C, Possnert G, Paabo S (1999) Nuclear DNA sequences from late Pleistocene megafauna. *Molecular Biology and Evolution* **16**, 1466-1473.
- Hackett SJ, Kimball RT, Reddy S, et al. (2008) A phylogenomic study of birds reveals their evolutionary history. *Science* **320**, 1763-1768.
- Haddrath O, Baker AJ (2001) Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proceedings of the Royal Society of London Series B-Biological Sciences* **268**, 939-945.
- Haile J, Holdaway R, Oliver K, et al. (2007) Ancient DNA chronology within sediment deposits: Are paleobiological reconstructions possible and is DNA leaching a factor? *Molecular Biology and Evolution* **24**, 982-989.
- Hamel J, Allen R, Davis L, et al. (2003) The Human Factor. In: *The Natural History of Southern New Zealand* (eds. Darby J, Fordyce RE, Mark A, Probert K, Townsend C). University of Otago Press, Dunedin, New Zealand.
- Hansen DM, Galetti M (2009) The Forgotten Megafauna. *Science* **324**, 42-43.
- Harshman J, Braun EL, Braun MJ, et al. (2008) Phylogenomic evidence for multiple losses of flight in ratite birds. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 13462-13467.
- Higuchi R, Bowman B, Freiburger M, Ryder OA, Wilson AC (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature* **312**, 282-284.
- Hofreiter M, Rabeder G, Jaenicke-Després V, et al. (2004) Evidence for Reproductive Isolation between Cave Bear Populations. *Current Biology* **14**, 40-43.
- Hofreiter M, Serre D, Poinar HN, Kuch M, Paabo S (2001) Ancient DNA. *Nature Reviews Genetics* **2**, 353-359.
- Holdaway RN, Jacomb C (2000a) Less is moa - Reply (vol 289, pg 1472, 2000). *Science* **290**, 53-53.
- Holdaway RN, Jacomb C (2000b) Rapid extinction of the moas (Aves : Dinornithiformes): Model, test, and implications. *Science* **287**, 2250-2254.
- Huynen L, Lissone I, Sawyer S, Lambert D (2008) Genetic identification of moa remains recovered from Tiniroto, Gisborne. *Journal of the Royal Society of New Zealand* **38**, 231-235.
- Huynen L, Millar CD, Scofield RP, Lambert DM (2003) Nuclear DNA sequences detect species limits in ancient moa. *Nature* **425**, 175-178.
- Johnson CN (2002) Determinants of loss of mammal species during the Late Quaternary 'megafauna' extinctions: life history and ecology, but not body size. *Proceedings of the Royal Society of London Series B-Biological Sciences* **269**, 2221-2227.
- Keepax CA (1981) Avian eggshell from archaeological sites. *Journal of Archaeological Science* **8**, 315-335.
- Lambert DM, Baker A, Huynen L, et al. (2005) Is a large-scale DNA-based inventory of ancient life possible? *Journal of Heredity* **96**, 279-284.

- McGlone MS, Wardle P, Worthy T (2003) Environmental change since the last glaciation. In: *The Natural History of Southern New Zealand* (eds. Darby J, Fordyce RE, Mark A, Probert K, Townsend C). University of Otago Press, Dunedin, New Zealand.
- McWethy DB, Whitlock C, Wilmshurst JM, McGlone MS, Li X (2009) Rapid deforestation of South Island, New Zealand, by early Polynesian fires. *Holocene* **19**, 883-897.
- Morin PA, Archer FI, Foote AD, et al. (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res* **20**, 908-916.
- Orlando L, Bonjean D, Bocherens H, et al. (2002) Ancient DNA and the Population Genetics of Cave Bears (*Ursus spelaeus*) Through Space and Time. *Mol Biol Evol* **19**, 1920-1933.
- Orlando L, Leonard JA, Thenot A, et al. (2003) Ancient DNA analysis reveals woolly rhino evolutionary relationships. *Molecular Phylogenetics and Evolution* **28**, 485-499.
- Oskam CL, Haile J, McLay E, et al. (2010) Fossil avian eggshell preserves ancient DNA. *Proceedings of the Royal Society B: Biological Sciences* **277**, 1991-2000.
- Owen R (1840) On the bone of an unknown struthious bird from New Zealand, meeting of November 12, 1839. *Proceedings of the Zoological Society London for 1839 Part VII*, 169-171.
- Paabo S (1985a) Molecular cloning of ancient Egyptian mummy DNA. *Nature* **314**, 644-645.
- Paabo S (1985b) Preservation of DNA in ancient Egyptian mummies. *Journal of Archaeological Science* **12**, 411-417.
- Paton T, Haddrath O, Baker AJ (2002) Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. *Proceedings of the Royal Society B: Biological Sciences* **269**, 839-846.
- Pereira SL, Baker AJ (2006) A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. *Molecular Biology and Evolution* **23**, 1731-1740.
- Phillips MJ, Gibb GC, Crimp EA, Penny D (2010) Tinamous and Moa Flock Together: Mitochondrial Genome Sequence Analysis Reveals Independent Losses of Flight among Ratites. *Systematic Biology* **59**, 90-107.
- Poinar HN, Hofreiter M, Spaulding WG, et al. (1998) Molecular coproscopy: Dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* **281**, 402-406.
- Poinar HN, Schwarz C, Qi J, et al. (2006) Metagenomics to paleogenomics: Large-scale sequencing of mammoth DNA. *Science* **311**, 392-394.
- Rasmussen M, Li YR, Lindgreen S, et al. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757-762.
- Rawlence NJ, Wood JR, Armstrong KN, Cooper A (2009) DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa. *Proceedings of the Royal Society B: Biological Sciences* **276**, 3395-3402.
- Scofield P, Worthy T, Holdaway RN, et al. (2005) Recent claims for more moa and huge errors in museum identifications - cutting through the spin. *Journal of Museums Aotearoa* **30**, 29-31.
- Shapiro B, Drummond AJ, Rambaut A, et al. (2004) Rise and fall of the Beringian steppe bison. *Science* **306**, 1561-1565.
- Slack KE, Delsuc F, McLenachan PA, Arnason U, Penny D (2007) Resolving the root of the avian mitogenomic tree by breaking up long branches. *Molecular Phylogenetics and Evolution* **42**, 1-13.
- Stiller M, Baryshnikov G, Bocherens H, et al. (2010) Withering away - 25,000 years of genetic decline preceded cave bear extinction. *Molecular Biology and Evolution* **27**, 975-978.
- Tennyson A, Martinson P (2006) *Extinct birds of New Zealand* Te Papa Press, Wellington.

- Tennyson AJD, Worthy TH, Jones CM, Scofield RP, Hand SJ (2010) Moa's Ark: Miocene Fossils Reveal the Great Antiquity of Moa (Aves: Dinornithiformes) in Zealandia. *Records of the Australian Museum* **62**, 105-114.
- Trewick SA, Gibb GC (2010) Vicars, tramps and assembly of the New Zealand avifauna: a review of molecular phylogenetic evidence. *Ibis* **152**, 226-253.
- Turvey ST, Green OR, Holdaway RN (2005) Cortical growth marks reveal extended juvenile development in New Zealand moa. *Nature* **435**, 940-943.
- Turvey ST, Holdaway RN (2005) Postnatal ontogeny, population structure, and extinction of the giant moa *Dinornis*. *Journal of Morphology* **265**, 70-86.
- Vickers-Rich P, Trusler P, Rowley MJ, et al. (1995) Morphology, myology, collagen and DNA of a mummified upland moa *Megalapteryx didinus* (Aves: Dinornithiformes) from New Zealand. *Tuhinga* **4**, 1-26.
- Wallis GP, Trewick SA (2009) New Zealand phylogeography: evolution on a small continent. *Molecular Ecology* **18**, 3548-3580.
- Willerslev E, Cooper A (2005) Ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **272**, 3-16.
- Willerslev E, Gilbert MTP, Binladen J, et al. (2009) Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. *BMC Evolutionary Biology* **9**.
- Willerslev E, Hansen AJ, Binladen J, et al. (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300**, 791-795.
- Wood JR, Rawlence NJ, Rogers GM, et al. (2008a) Coprolite deposits reveal the diet and ecology of the extinct New Zealand megaherbivore moa (Aves, Dinornithiformes). *Quaternary Science Reviews* **27**, 2593-2602.
- Wood JR, Worthy TH, Rawlence NJ, Jones SM, Read SE (2008b) A deposition mechanism for Holocene miring bone deposits, South Island, New Zealand. *Journal of Taphonomy* **6**, 1-20.
- Worthy TH (1987) Sexual dimorphism and temporal variation in the North Island moa species *Euryapteryx curtus* (Owen) and *Pachyornis mappini* (Archy). *National Museum of New Zealand Records* **3**, 59-70.
- Worthy TH (2007a) Moas and phylogenomics: How nomenclatural errors do a disservice to the understanding of moa taxonomy. *Auk* **124**, 1447-1449.
- Worthy TH (2007b) Moas and phylogenomics: How nomenclatural errors do a disservice to the understanding of moa taxonomy. *Auk* **124**, 1447-1449.
- Worthy TH, Holdaway RN (1996) Quaternary fossil faunas, overlapping taphonomies, and palaeofaunal reconstruction in north Canterbury, South Island, New Zealand. *Journal of the Royal Society of New Zealand* **26**, 275-361.
- Worthy TH, Holdaway RN (2002) *The lost world of the Moa* Canterbury University Press, Christchurch.

Aims and scopes of the thesis

This project is based on aDNA extractions from 290 individuals and 217 radiocarbon dates representing four moa species, excavated from Holocene fossil deposits in North Canterbury, South Island, New Zealand. By analysing DNA from many fossils preserved in close spatial and (relatively close) temporal proximity, it should be possible to interpret the results in the context of populations. Such studies are rare in aDNA research and this project represents the first attempt to study moa at the population level. The overall aim of this thesis is to *apply aDNA technology to gain new insights on the population structure and biology of four sympatric moa populations, as they appeared prior to their extinction*. In other words: a detailed genetic reflection of New Zealand's extinct megafauna in their last era. The following serves as an overview of the aims of this project, grouped into four main categories. More careful elaborations on the background and theory accompanying these objectives are found in the relevant chapters.

1.2.1 Aims and scopes

(1) To obtain a molecular species and sex identification for all sampled individuals

This aim can be seen as the 'basic genetic component' of my research. By applying appropriate molecular methods, I will attempt to establish the taxon and sex of each sampled moa. This task is trivial when studying extant species but a considerable challenge when studying extinct ones. Having secure species and sex ID's are crucial in all further analyses of genetic diversity, and not least, in the stable isotope research (presented elsewhere) that runs in parallel with the genetics. We need to know exactly what we are studying. Also, from these basic genetic components I hope to obtain new information on the species and sex composition of the North Canterbury moa assemblages. Some previous efforts have indicated that sex ratios in moa were skewed towards an excess of females (e.g., Bunce et al., 2003; Huynen et al., 2003; Holdaway, Worthy, 1996), but these were only partially based on genetic identifications and/or covered vast geographic and temporal distances, and hence unlikely to represent the living moa assemblages. By assessing the sex ratios in both juveniles and adults from the four North Canterbury moa species, I hope to be able to discuss certain aspects of moa biology such as, sex related mortality, and ecological niche partitioning. Taphonomic aspects can also be investigated in this frame work, including an assessment of depositional biases among sex and species, as well as the concurrency between morphologically-determined and molecular taxon

identifications, which has previously been the subject of some controversy (Baker et al., 2005; Scofield et al., 2005).

(2) To apply molecular technology to characterise a newly-discovered fossil site

Whilst working on this project, a North Canterbury fossil deposit was discovered (Rosslea) close to the two main sites (Pyramid Valley and Bell Hill Vineyard). An inclusion of newly excavated moa fossils from Rosslea offers some interesting possibilities. The aim in this context is first of all to scientifically describe this important new site but rather than a traditional paleontological assessment, the characterisation will have a strong molecular focus. In addition to provide a secure molecular species identification of each Rosslea bone, the aim is also to assess the level of DNA preservation. This will be done in a comparative investigation involving qPCR and hundreds of bones from the three sites. Radiocarbon ages will be included to investigate aDNA degradation through time. I am particularly interested in investigating whether long term DNA decay in bone follows a model of first order decay kinetics as has been proposed previously (e.g., Lindahl 1993) but never shown empirically across a considerable timeframe. Such assessment can potentially provide significant new insights on the rate and mode of post mortem DNA degradation, beneficial to the field of aDNA as well as other disciplines dealing with degraded DNA (i.e. forensic sciences).

(3) To apply high-throughput sequencing (HTS) in an attempt to identify microsatellites from aDNA

Highly variable microsatellite loci offer a powerful approach to study population genetics, and would constitute a highly informative ‘nuclear alternative’ to the mtDNA sequences that are commonly analysed in aDNA contexts. Realising the potential in a moa microsatellite study, HTS will be applied in an attempt to identify these repetitive sequences in the nuclear moa genome. If successful, this method could prove advantageous not only in an aDNA context but also to other research areas where microsatellite markers are being applied. However, an identification of these repetitive regions is just the first step towards a high resolution population genetic study of moa. Primers have to be developed, optimised and tested for genetic variation. Moreover, there can be numerous problems associated with the amplification of degraded DNA and low copy numbers. Hence a careful and systematic evaluation of the results and caveats is necessary to establish the protocols ensuring sufficient data quality. A considerable proportion of the thesis is dedicated to this process.

(4) To study moa genetics at the population level

Given that the generation of microsatellite data is successful, a whole series of genetic analyses can be carried out based on both mtDNA and microsatellite data. The four sympatric moa populations can be examined to gain new insights on the population dynamics and biology of these extinct ratites. New Zealand was the last major landmass to be occupied by humans and genetic data can, when combined with the ^{14}C ages, provide a detailed description of megafauna gene pools in a pristine environment.

More specifically, the three genetic components (sex ID, mtDNA and microsatellite data) can be used to investigate a range of topics from basic descriptive measures of genetic diversity in the four sympatric species, to more advanced analyses of the demography and genetic stability and structure through time. For example, it can be assessed, whether the moa gene pools carried any signals of the hypothesised late Holocene population collapse (Gemmell et al., 2004). A massive decline in numbers should be accompanied by evidence of substantial genetic drift and loss of genetic diversity through time. This can be tested with the high level of resolution offered by microsatellite data. Also, kinship and sex biased dispersal patterns can potentially be analysed. For example, the presence of closely related individuals among fossils excavated from the same site could perhaps shed light on the social structure in a moa population.

In short, the objective with these population genetic analyses is two-fold. First, they should provide genetic insights at an unprecedented level in aDNA research, and add to our general understanding of moa biology and the pre-human New Zealand ecosystem. Secondly, the analyses can serve as a thorough exploration of novel data in unknown analytical territory. It is not known to what extent the commonly applied allele-frequency based methods can apply to microsatellite data covering 4000 years. With a combination of radiocarbon ages and microsatellite data, the potential to capture micro-evolutionary patterns in real-time seems unprecedented, but it is unknown what kind of information can be extracted from such heterochronous microsatellite data. This thesis aims to explore this potential.

1.2.2 References

- Baker AJ, Huynen LJ, Haddrath O, Millar CD, Lambert DM (2005) Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: The giant moas of New Zealand. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8257-8262.
- Bunce M, Worthy TH, Ford T, et al. (2003) Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425**, 172-175.

- Gemmell NJ, Schwartz MK, Robertson BC (2004) Moa were many. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, 430-432.
- Holdaway RN, Worthy TH (1997) A reappraisal of the late quaternary fossil vertebrates of Pyramid Valley Swamp, North Canterbury, New Zealand. *New Zealand Journal of Zoology* **24**, 69-121.
- Huynen L, Millar CD, Scofield RP, Lambert DM (2003) Nuclear DNA sequences detect species limits in ancient moa. *Nature* **425**, 175-178.
- Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715.
- Scofield P, Worthy T, Holdaway RN, et al. (2005) Recent claims for more moa and huge errors in museum identifications - cutting through the spin. *Journal of Museums Aotearoa* **30**, 29-31.

The North Canterbury moa

The 290 moa fossils sampled for this thesis are all of Holocene age (except for one *Dinornis robustus* bone from Glencrieff, see Chapter 4). All bones have been excavated from a relatively small area near Waikari in North Canterbury, South Island, New Zealand (Figure 1.3.1). This paleontological setting allowed a large, cohesive investigation of moa population genetics, rather than a study of small disconnected cases, as is typical in aDNA research. This section introduces the four moa species dealt with in this thesis, and provides a brief description of the North Canterbury Holocene environment, as it appeared prior to human arrival in New Zealand. It provides an ecological context in which the genetic data are analysed and interpreted in the remaining chapters.

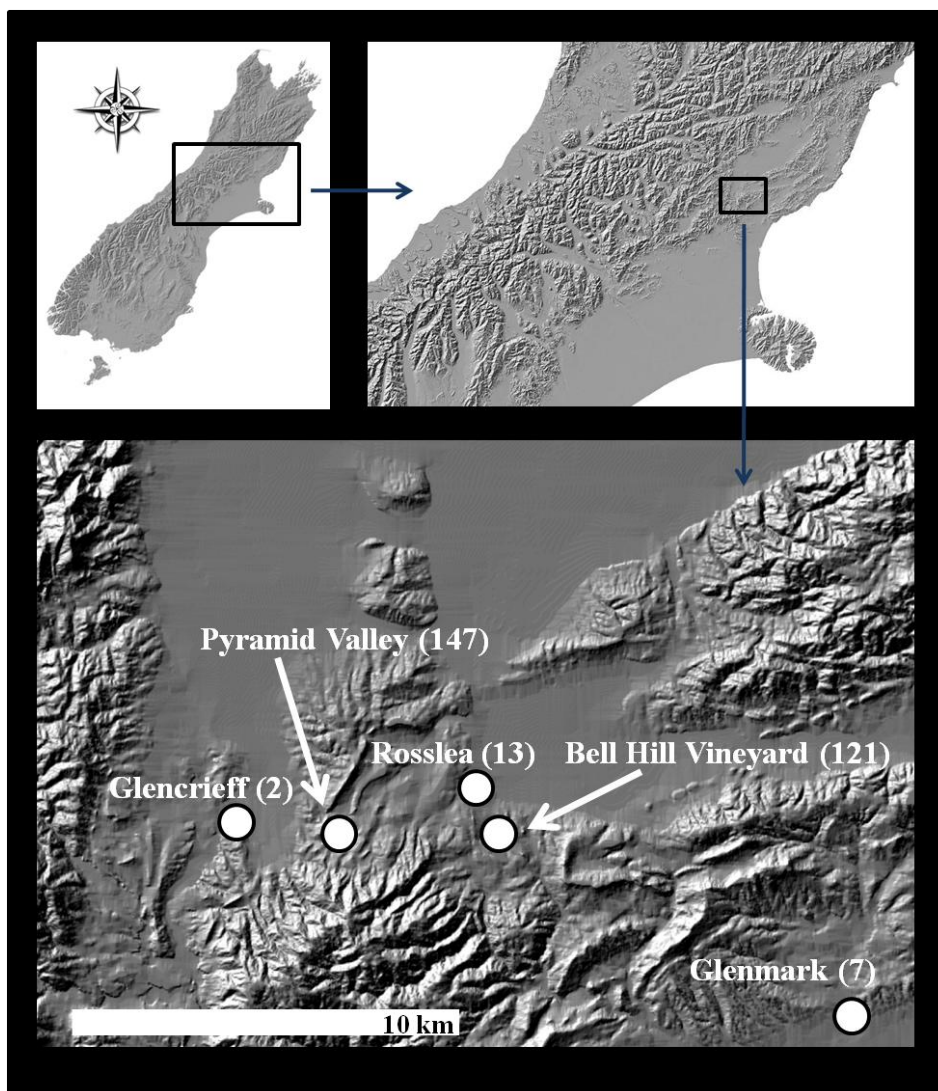


Figure 1.3.1: The South Island of New Zealand with the study area in North Canterbury highlighted. The 290 sampled moa fossils included in this study have been excavated from five deposits. The number of sampled bones is indicated for each site.

Six species of moa are known from deposits of Holocene age in North Canterbury. Two of them (*Megalapteryx didinus* and *Anomalopteryx didiformes*) were very rare in the Waikari region, whereas the other four species existed sympatrically for thousands of years before disappearing suddenly 500-600 years ago (Worthy, Holdaway, 1996). Unless stated otherwise, the information in the four species descriptions below is from Worthy and Holdaway (2002) or Tennyson and Martinson (2006).

1.3.1 *Dinornis robustus* (South Island giant moa)

This was the largest moa species and one of the largest birds ever, probably surpassed in weight only by *Aepyornis maximus* (Madagascan elephant birds), and the Miocene Australian dromornithids (Murray, Vickers-Rich, 2004). An adult female weighed up to 242 kg, whereas the males were considerably smaller (see Figure 1.1.5), at up to 84 kg. *Dinornis* fossils have been found in a range of habitats including forests and open scrubland, and even in sub-alpine caves, which suggests that it was a generalist. The species is common in North Canterbury Holocene deposits. One of the most spectacular paleontological finds in New Zealand is arguable the large number of more or less complete and well-preserved *Dinornis* skeletons excavated from the famous Pyramid Valley fossil deposit (Holdaway, Worthy, 1997; Worthy, Holdaway, 1996).

1.3.2 *Euryapteryx curtus* (Stout-legged moa)

This emeid was a robust moa with short legs. Although *Euryapteryx* did not show North-South Island monophyly based on genetic data (Figure 1.1.3), it exhibited an extreme variability in size. The North Canterbury *Euryapteryx* were large animals, weighing as much as an ostrich (app. 80 kg), whereas some North Island individuals were about the size of a turkey, weighing only 12-15 kg. Fossils of *Euryapteryx curtus* have been found in many different environments, including the coastal regions, which could perhaps explain how this species have apparently maintained gene flow between the two main islands longer than the other species. Of other notable features, *Euryapteryx* had an unusually long windpipe, suggesting that it was able to produce loud calls. It is frequently encountered in archaeological sites.

1.3.3 *Pachyornis elephantopus* (Heavy-footed moa)

This moa had very robust legs and was the largest of the Emeids (Figure 1.3.2). The sampling of >30 *Pachyornis* tibiotarsi for this study confirmed that the bones were impressively dense and heavy, and it took a considerable effort to drill out a sample. Although the species was frequent in the region, it

was the rarest of the four common North Canterbury moa in the Holocene (see Chapter 2, Section 1). The mean weight was estimated to be 87 kg, but again, the females were significantly larger than the males (Allentoft, unpublished data). Rawlence et al. (2009) identified moa feathers using DNA, and showed that *Pachyornis elephantopus* had both a white-tipped and black-tipped feathers. This feature would present a speckled appearance as seen in, for example, the great spotted kiwi (*Apteryx haastii*). Two *Pachyornis* species were present on the South Island (Figure 1.1.4) but fossils of *Pachyornis australis* (crested moa) have been found at higher altitudes and further South than *P. elephantopus*. *Pachyornis australis* seems to have been more abundant during the cooler periods in the Pleistocene.

Because of the copyright, the image cannot be shown in the online (digital) version of this thesis.

Figure 1.3.2: Reconstruction of a *Pachyornis elephantopus* crossing a stony riverbed. This species had massive legs and was the largest member of Emeidae.

Image reference: Tennyson and Martinson (2006).

1.3.3 *Emeus crassus* (Eastern moa)

The Eastern moa was the smallest of the four common moa in the district, having a more slender build than either *Euryapteryx curtus* or *Pachyornis elephantopus* and it was much smaller than *Dinornis*. It was common on the eastern South Island and abundant in the North Canterbury Holocene environment. Adults weighed between 36-79 kg, and sexual size dimorphism was evident in this species as well (Allentoft, unpublished data). A relatively delicate beak and fruit remains preserved in gizzards suggest that this species had a less fibrous diet than other moa species.

1.3.4 The North Canterbury environment

The Waikari region, at the southern ends of the tectonically active Hawarden Basin has an environment typical for the eastern rain shadows of the Southern Alps. Most of the area is >300 meters above sea level and consists of flat alluvial plains and rolling downlands (Figure 1.3.3).

Drought is common in the area, partly facilitated by hot, dry north-westerly föhn winds in summer (Worthy, Holdaway, 1996). The annual rainfall is 600-900 mm per year (Sturman, 2008). Most of the eastern forests of the South Island were destroyed by anthropogenic fires shortly after humans settled in the late 13th century (McWethy et al., 2009). Today the area supports intensive pastoral farming and vineyards, largely deprived of the original vegetation.



Figure 1.3.3: A typical North Canterbury landscape. The shallow seasonal lake covers the Pyramid Valley deposit, situated between rolling limestone hills. Before human settlement in New Zealand, the area was covered by a mosaic of forest and shrubland.

Photo: Richard N. Holdaway

The biota of New Zealand's South Island has been shaped by recurring glacial cycles throughout the Pleistocene, and the environment of the Waikari region also changed considerably during the period represented in the fossil record dealt with in this thesis (c. 13,000-600 years BP).

During the last glacial maximum (LGM) (c. 25,000-15,000 years BP), the average temperatures were 4-6 °C lower than today, and large glaciers covered much of the high country to the West. After a 5000 year transition phase (late glacial warming), the Holocene (past 10,000 years) were characterised by temperatures close to present day (McGlone et al., 2003). Pollen records from sites throughout the central South Island (Burrows, Russell, 1990; McGlone et al., 1997; McGlone et al., 2004; Moar, 1971; Moar, 1973) suggest a non-synchronous but repeated pattern of establishment of post-glacial shrubland (14,000-10,000 years BP), followed by spread of tall podocarp forest (13,600–7500 years BP), and an apparent later expansion of *Nothofagus* dominated forest (>7500–2000 years BP) in the high country of North Canterbury. Although most of the region was clothed in woody vegetation in the late Holocene, a mosaic of forest and shrubland would have offered a heterogeneous array of habitats (e.g., Batcheler, 1989; Burrows, 1989), sustaining the four sympatric moa species.

1.3.5 The sites

New Zealand has one of the richest late Quaternary avifaunal records in the world (Worthy, Holdaway, 2002), and the area around Waikari is arguably one of the most important regions because of its diverse fossil fauna which accumulated concurrently in many sites. Several sites have yielded large collections of well-preserved bones, providing ample material for investigating patterns of taphonomy (Wood et al., 2008; Worthy, Holdaway, 1996) and genetics. The area has a long history of paleontological finds, starting with the discovery of Glenmark Swamp in 1866 and the excavation in the following years by Julius von Haast. Thousands of moa bones were uncovered and many were sent abroad in exchange for display material for the nascent Canterbury Museum (Worthy, Holdaway, 1996). Later, the extensive Pyramid Valley excavations in the mid 20th century proved highly important because the many moa were preserved as complete or semi-complete skeletons allowing discrimination between individuals.

Although the deposits in the area include caves, predation sites and wetlands, the work reported here is based exclusively on bones that have been excavated from wetlands. Fossils from five different sites were sampled for ancient DNA (Figure 1.3.1). Pyramid Valley (PV) (Figure 1.3.3) is a lacustrine deposit, whereas Bell Hill Vineyard (BHV), Rosslea, Glenmark and Glencrieff are all spring-holes where birds got preserved in peats or clays (see Wood et al., 2008 and Chapter 2 in this thesis). The sediments in these deposits differ, but the bones were typically preserved in organic-rich mud or peat, or calcareous “gyttja” at PV (Gregg, 1972), buffered from limestones in the area (Wood et al., 2008; Worthy, Holdaway, 1996). The carbonate buffering of otherwise acidic peats probably explains the remarkable bone preservation that characterises the area (see Chapter 2).

Pyramid Valley (42°58'22.0"S, 172°35'49.0"E) and Bell Hill Vineyard (42°58'19.36"S, 172°39'56.15"E) contributed most of samples in this study (268 bones, representing the same number of individuals). The two sites are described in Chapter 2 (Section 1). Rosslea (42°57'53.83"S, 172°39'22.39"E) is a newly-discovered deposit that was exposed by a local landowner in 2008 during earthworks and 13 moa bones were sampled for this research. Chapter 2 (Section 2) includes the first description of Rosslea and a list of the material recovered from that site. Bones from Glenmark and Glencrieff were sampled late in the research programme, in an attempt to increase the temporal depth of the genetic data. These two sites are described briefly below. Information on the samples and radiocarbon dates mentioned here is given in Digital Appendix 2. The methodology is described in Chapter 2 (Section 2), and the temporal genetic analyses involving radiocarbon dates are in Chapters 2, 4 and 5.

Glenmark

The Glenmark deposit (43°00'00.0"S, 172°46'50.0"E) is 10 km southeast of Waikari and was the first major New Zealand Quaternary deposit to be exploited. The sediments contained moa bones in high concentrations down to a depth of 4 m (Anderson, 1989). The early excavation yielded many significant finds, including the largest moa bone ever found (Figure 1.3.4), allegedly described in Owen (1869) (Paul Scofield, pers. comm.), and the type specimens of Haast's eagle (*Hieraaetus moorei*) (Holdaway, 1994). An exhaustive study of the fauna from this site has not been conducted because most of the specimens are scattered, sent to museums and other collections around the world. However, a crude assessment of the fauna was conducted in Worthy (1990) who converted Haast's original list into the moa taxonomy of 1990. He concluded that there were 57% *Emeus crassus*, 15% *Pachyornis elephantopus*, 12% *Euryapteryx geranoides* (now *E. curtus*), 8% *Dinornis struthoides*, 2% *D. novaezealandiae*, and 5.5% *D. giganteus*. The latter three taxa all refer to *Dinornis robustus* which therefore constituted c. 15% of the local fauna. Bones of two individuals of *Megalapteryx didinus* (upland moa) and two *Anomalopteryx didiformes* (little bush moa) were also found in the deposit. Both were rare in this area (Worthy, Holdaway, 1996). Three radiocarbon dates on moa remains were available for Glenmark before this research, 2730 ± 70 BP and 7110 ± 109 BP (Anderson, 1989) and 2302 ± 31 BP (Bunce et al., 2009) respective, suggesting at least 5000 years of Holocene fossil accumulation. Eight moa bones from Glenmark were radiocarbon dated for this project (Chapter 4).



Figure 1.3.4: A 991 mm tibiotarsus from a south Island giant moa (*Dinornis robustus*), recovered from Glenmark Swamp in 1868. The largest moa bone ever found (described in Owen, 1869).

Photo: Malene Møhl.

Glencrieff

The Glencrieff site (42°58'07.45"S, 172°34'01.84"E) is an isolated spring-hole on a pasture 1.6 km southwest of PV. The site has been excavated on several occasions (most recently in 2007; Rawlence et al. in prep.) since its discovery in 1971 (Wood et al., 2008). Moa were trapped in the soft sediments and their bones accumulated and were preserved in an anoxic peat. Worthy and Holdaway (1996)

provided a full account of the history, stratigraphy, and taphonomy of Glencrieff, recording hundreds of bones of *Pachyornis elephantopus*, *Emeus crassus* and *Dinornis robustus*. Interestingly, the last of the ‘eastern four’, *Euryapteryx curtus*, has not been found in this site. Worthy and Holdaway (1996) hypothesised that this species preferred well-drained soils and its absence supports the view that moa taxa had clearly defined niches (see Chapter 2, Section 1). Three bones from Glencrieff had been radiocarbon dated before this research, 11898 ± 82 BP; 10470 ± 130 BP and 10480 ± 120 BP (Worthy, Holdaway, 1996). These dates are considerably older than the mid- to late Holocene dates of most sites in this region.

A large proportion of the Glencrieff bones have now been lost to a private collection overseas (Wood et al., 2008) making the bulk of material from this site inaccessible to further research. This is extremely unfortunate in light of its greater age, which includes the important Pleistocene-Holocene transition phase. Two Glencrieff bones (*Dinornis robustus*) were sampled and radiocarbon dated for this research (see Chapter 4).

1.3.6 References

- Anderson A (1989) *Prodigious Birds* Cambridge University Press, Cambridge, England.
- Batcheler CL (1989) Moa browsing and vegetation formations, with particular reference to deciduous and poisonous plants *New Zealand Journal of Ecology* **12**, 57-65.
- Bunce M, Worthy TH, Phillips MJ, et al. (2009) The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20646-20651.
- Burrows CJ (1989) Moa browsing - evidence from the Pyramid Valley mire. *New Zealand Journal of Ecology* **12**, 51-56.
- Burrows CJ, Russell JB (1990) Aranuiian vegetation history of the Arrowsmith Range, Canterbury.1. Pollen diagrams, plant macrofossils, and buried soils from Prospect Hill. *New Zealand Journal of Botany* **28**, 323-345.
- Gregg D (1972) Holocene stratigraphy and moas at Pyramid Valley, North Canterbury, New Zealand. *Records of the Canterbury Museum* **9**, 151-158.
- Holdaway RN (1994) An Exploratory Phylogenetic Analysis of the Genera of the Accipitridae, with Notes on the Biogeography of the Family in B.-U. Meyburg, and R. D. Chancellor, eds. *Raptor Conservation Today*, Berlin: World Working Group on Birds of Prey and Owls and the Pica Press, 601-649.
- Holdaway RN, Worthy TH (1997) A reappraisal of the late quaternary fossil vertebrates of Pyramid Valley Swamp, North Canterbury, New Zealand. *New Zealand Journal of Zoology* **24**, 69-121.

- McGlone MS, Moar NT, Meurk CD (1997) Growth and vegetation history of alpine mires on the Old Man Range, Central Otago, New Zealand. *Arctic and Alpine Research* **29**, 32-44.
- McGlone MS, Turney CSM, Wilmshurst JM (2004) Late-glacial and Holocene vegetation and climatic history of the Cass basin, central south island, New Zealand. *Quaternary Research* **62**, 267-279.
- McGlone MS, Wardle P, Worthy T (2003) Environmental change since the last glaciation. In: *The Natural History of Southern New Zealand* (eds. Darby J, Fordyce RE, Mark A, Probert K, Townsend C). University of Otago Press, Dunedin, New Zealand.
- McWethy DB, Whitlock C, Wilmshurst JM, McGlone MS, Li X (2009) Rapid deforestation of South Island, New Zealand, by early Polynesian fires. *Holocene* **19**, 883-897.
- Moar NT (1971) Contributions to the Quaternary history of the New Zealand flora: 6. Aranuian pollen diagrams from Canterbury, Nelson, and North Westland, South Island. *New Zealand Journal of Botany* **9**, 80-145.
- Moar NT (1973) Contributions to the Quaternary history of the New Zealand flora: 7. Two Aranuian pollen diagrams from central South Island. *New Zealand Journal of Botany* **11**, 291-304.
- Murray PF, Vickers-Rich P (2004) *Magnificent mihirungs: the colossal flightless birds of the Australian dreamtime* Indiana University Press, Bloomington.
- Owen R (1869) On *Dinornis* (Part XII): Containing a description of the femur, tibia, and metatarsus of *Diniornis maximus*. *Transactions of the Zoological Society of London* **6**, 497-500.
- Rawlence NJ, Wood JR, Armstrong KN, Cooper A (2009) DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa. *Proceedings of the Royal Society B: Biological Sciences* **276**, 3395-3402.
- Sturman A (2008) Weather and climate. In: *Natural history of Canterbury* (eds. Winterbourn MJ, Knox GA, Burrows CJ, Marsden ID). Canterbury University Press, Christchurch, New Zealand.
- Tennyson A, Martinson P (2006) *Extinct birds of New Zealand* Te Papa Press, Wellington.
- Wood JR, Worthy TH, Rawlence NJ, Jones SM, Read SE (2008) A deposition mechanism for Holocene miring bone deposits, South Island, New Zealand. *Journal of Taphonomy* **6**, 1-20.
- Worthy TH (1990) An Analysis of the Distribution and Relative Abundance of Moa Species (Aves, Dinornithiformes). *New Zealand Journal of Zoology* **17**, 213-241.
- Worthy TH, Holdaway RN (1996) Quaternary fossil faunas, overlapping taphonomies, and palaeofaunal reconstruction in north Canterbury, South Island, New Zealand. *Journal of the Royal Society of New Zealand* **26**, 275-361.
- Worthy TH, Holdaway RN (2002) *The lost world of the Moa*. Canterbury University Press, Christchurch.

Chapter 2: Basic genetic components: taxon, sex and preservation

Introducing the manuscripts in Chapter 2

This chapter includes two manuscripts. The first “Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand’s extinct megafauna”, was published earlier this year [Quaternary Science Reviews 29 (2010) 753-762]. The published paper is found in Appendix 3 and the electronic version in Digital Appendix 18.

This manuscript was based on the basic genetic components of the project, namely species and sex identifications. With that data at hand, the fundamental genetic aims of the research programme were accomplished. However, the level of information that could be extracted directly from this basic information was beyond expectations. Having samples from four species and several sites offered some excellent comparative opportunities and highly significant signals among species and fossil sites were revealed. This prompted the novel interpretations on moa biology and behaviour presented in the manuscript. Being able to assess biology of extinct taxa on aspects like sex ratios and territorial behaviour shows the level of resolution that can be achieved when obtaining hundreds of samples from a small confined area. In addition to the new information on moa, the analyses documented that fossil deposits, proximate in space and time, recorded the surrounding Holocene environments very differently. It became clear that each site represents an independent window into the past, and that different site-specific taphonomic biases prevail. Hence, several variables are complicating the observed patterns and despite simple and significant results, the interpretations are not straightforward.

The second (unpublished) manuscript relies on the same basic molecular tools as above to describe a newly discovered fossil site (Rosslea) in North Canterbury. While this task was never visualised as part of the original research programme, the finding of new fossils in the area of focus provided an opportunity to add new information to the picture. The manuscript shows how aDNA technology can work as “applied science” by assisting in the description of fossils material, but the work extends beyond basic genetic species identifications to also characterise DNA preservation in the bones. The view that biomolecule preservation can be an important taphonomic character is advocated, in that descriptions of Pleistocene and Holocene fossil sites will benefit greatly from including the molecular element. Lastly, DNA preservation is examined in a temporal context. By incorporating many radiocarbon ages and data from several sites, a comprehensive investigation of post mortem DNA degradation through time, is carried out.

My role has been to obtain and process the samples, to conduct the laboratory work and the analyses, and finally to conceptualise and write the manuscripts. All co-authors contributed to the writing and the interpretations of the published paper.

Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand's extinct megafauna

Morten E. Allentoft^{1,2}, Michael Bunce², R. Paul Scofield³, Marie L. Hale¹ and Richard N. Holdaway^{1,4}

1) School of Biological Sciences, University of Canterbury, Private bag 4800, Christchurch, New Zealand.

2) Ancient DNA laboratory, School of Biological Sciences and Biotechnology, Murdoch University, Perth, Australia.

3) Canterbury Museum, Rolleston Ave, Christchurch 8013, New Zealand.

4) Palaeocol Research Ltd, P.O. Box 16 569, Hornby, Christchurch 8042, New Zealand.

2.1.1 Abstract

Ancient DNA was isolated from the bones of 267 individuals of the extinct New Zealand moa (Aves: Dinornithiformes) from two late Holocene deposits [Pyramid Valley (PV) and Bell Hill Vineyard (BHV)] located 5.7 km apart in North Canterbury, South Island. The two sites' combined fossil record cover the last 3000 years of pre-human New Zealand and mitochondrial DNA confirmed that four species (*Dinornis robustus*, *Euryapteryx curtus*, *Emeus crassus*, and *Pachyornis elephantopus*) were sympatric in the region. However, the relative species compositions in the two deposits differed significantly with *D. robustus* and *E. crassus* being most abundant at PV while *E. curtus* outnumbered the other three moa taxa combined at BHV. A subsample of 227 individuals had sufficient nuclear DNA preservation to warrant the use of molecular sexing techniques, and the analyses uncovered a remarkable excess of females in both deposits with an overall male to female ratio of 1 : 5.1. Among juveniles of *E. curtus*, the only species which was represented by a substantial fraction of juveniles, the sex ratio was not skewed (10 ♂, 10 ♀), suggesting that the observed imbalance arose as a result of differential mortality during maturation. Surprisingly, sex ratios proved significantly different between sites with a 1 : 2.2 ratio at BHV ($n = 90$) and an entire 1 : 14.2 at PV ($n = 137$). Given the mobility of large ratites, and the proximity of the two fossil assemblages in space and time, these differences in taxonomic and gender composition indicate that moa biology and the local environment have affected the fossil representation dramatically and several possible explanations are offered. Apart from adding to our understanding of moa biology, these discoveries reinforce the need for caution when basing interpretation of the fossil record on material from a single site.

2.1.2 Introduction

Moa (Aves: Dinornithiformes) occupied most areas of New Zealand from the coastline to the inland alpine areas until they disappeared shortly after the arrival of human settlers in the late 13th century CE (e.g. Holdaway, Jacomb, 2000; Worthy, Holdaway, 2002). Since the early-1990s, ancient DNA (aDNA) technology has been applied to investigate the systematics and palaeobiology of these extinct ratites and has shed light on aspects of moa evolution (e.g., Baker et al., 2005; Cooper et al., 2001; Cooper et al., 1992; Haddrath, Baker, 2001) their diet (Wood et al., 2008), and plumage (Rawlence et al., 2009). aDNA analyses also provided the basis for arguably the most significant finding in recent moa research, when it was revealed that highly divergent moa allomorphs were actually single species displaying extreme levels of reverse sexual size-dimorphism (RSD) (Bunce et al., 2003; Huynen et al., 2003). As a result, moa taxonomy was revised, emphasizing the advantages in including a genetic approach in phylogenetic, taxonomic, and palaeoecological analyses.

The present study represents the next natural step forward in moa research, shifting focus from larger evolutionary patterns towards population level dynamics. By sampling many individuals preserved in close spatial and temporal proximity, it should be possible to study aspects of moa biology not accessible in fossils covering many thousands of years and large geographical distances. The late Holocene environment of North Canterbury, South Island, New Zealand (Figure 2.1.1) supported a diverse moa fauna largely dominated by four of the ten presently-recognized moa species (*Dinornis robustus*, *Emeus crassus*, *Euryapteryx curtus*, *Pachyornis elephantopus*) and is ideal for such research because of several excavated fossil-rich deposits in this region (Worthy, Holdaway, 1996).

Museum collections from two adjacent sites, Pyramid Valley (PV) and Bell Hill Vineyard (BHV) in North Canterbury, include >300 individual well-preserved, non-mineralised fossils of partial and complete moa skeletons, providing an opportunity to study the moa at the population level. A few moa from PV had been radiocarbon dated and they suggested a late Holocene fossil accumulation: conventional radiocarbon ages spanned the period from 3740 BP to 1245 BP (Blau et al., 1953; Burrows et al., 1981; Gregg, 1972; Kulp et al., 1952). This rudimentary chronology was supported by four radiocarbon ages of sediment layers surrounding the moa remains (Gregg, 1972). Large temporal overlaps among species and a largely contemporaneous deposition at the two sites ranging from c. 3550 B.P. to c. 700 B.P. were confirmed by a suite of >110 radiocarbon ages of moa from both sites and all four species (Holdaway et al. in prep.). The relatively narrow temporal window allows us to interpret the data in the context of moa populations, as they appeared in their last era, just before humans reached New Zealand.

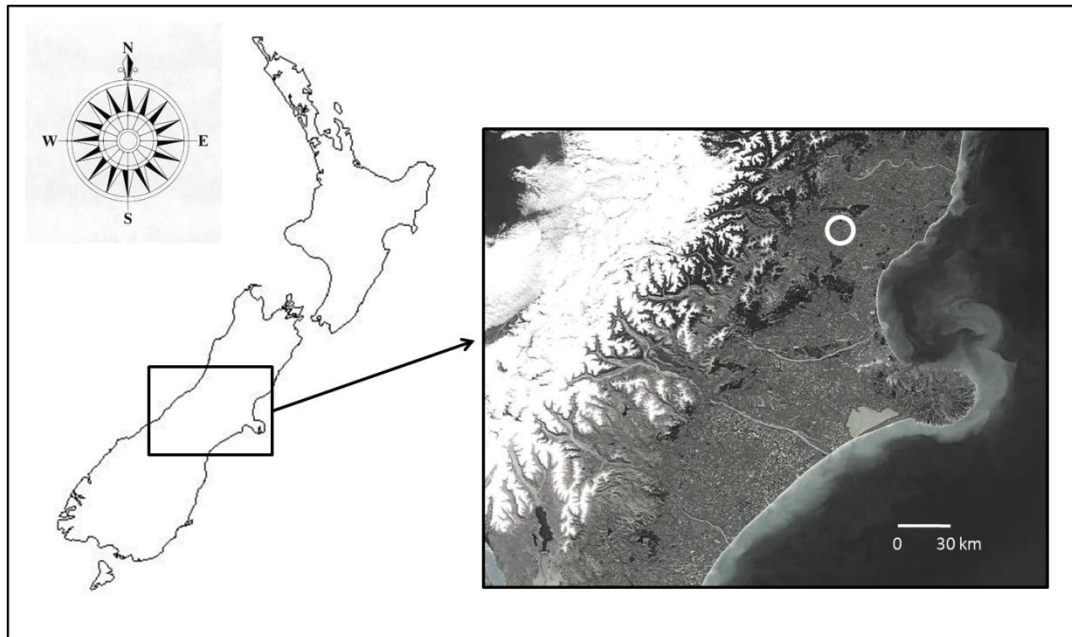


Figure 2.1.1: Map of New Zealand with the central South Island highlighted. The white circle frames the locations of Pyramid Valley (PV) and Bell Hill Vineyard (BHV), near Waikari in North Canterbury. See Digital Appendix 3 for relative locations of sites. Satellite image was downloaded and modified (with permission) from the European Space Agency (www.esa.int).

Population-based studies of extinct species are unusual, because fossils with good biomolecule preservation are scarce. Although aDNA was used to assess overall patterns of population dynamics within cave bear (*Ursus spelaeus*) (Hofreiter et al., 2004; Orlando et al., 2002), woolly mammoth (*Mammuthus primigenius*) (Barnes et al., 2007; Debruyne et al., 2008; Gilbert et al., 2008), Berengian steppe bison (*Bison priscus*) (Shapiro et al., 2004) and various extinct lions (*Panthera leo* ssp.) (Barnett et al., 2009), none of these discussed local scale population biology, as is here attempted with the North Canterbury moa.

We extracted and amplified aDNA to identify 267 individuals to species level and establish their relative abundance at each site. The high species diversity reported within moa [nine species presently recognized (Bunce et al., 2009), many of them sympatric] and the demonstrated sexual dimorphism, combined with possible intra-species size variation among fossils from different geographic regions and climatic eras (Worthy, 1987; Worthy, Holdaway, 1996), have resulted in as many as 64 moa species and 20 genera being proposed since the 1840s (Worthy, Holdaway, 2002). Exacerbating the taxonomic minefield, damaged bones and juvenile bones are often not identifiable to species at all by traditional morphological and morphometric approaches. For example, 28 of the sampled leg bones from BHV were catalogued simply as “juvenile emeids” (i.e. indeterminate members of the family “Emeidae” within moa). A DNA-based identification methodology is therefore the crucial first step in assessing accurately the species composition of the moa fauna in each deposit. Establishing the

relative abundances at such a fine geographic scale can provide fundamental knowledge on the complexity characterising the living moa assemblage. Not least, an accurate species composition also contributes significantly to the interpretation of the deposition modes, including potential size-specific effects, given, for example, that females of *D. robustus* were up to 2 m high at the back weighing up to 250 kg (Figure 2.1.2), whereas an average adult of *E. crassus* was c. 1 m high and weighed <60 kg (Worthy, Holdaway, 2002).

Finally, by targeting specific DNA sequences on the sex chromosomes, we could determine the gender of each moa individual and hence the sex ratios within each species, as represented in the two sites. The governing sex ratio can reveal much about the reproductive biology and evolutionary history of a species (Hardy, 1997; Hardy, 2002; Sheldon, 1998; West et al., 2002) and seems particularly interesting as moa display the largest reported reverse sexual dimorphism known in any terrestrial vertebrate (Figure 2.1.2). Previous work has suggested that moa sex ratios were somewhat unbalanced. Huynen et al. (2003) genetically sexed 115 moa of various taxa from different geologic ages and locations throughout New Zealand and reported an overall female-biased sex ratio of 1 : 1.7. Bunce et al. (2003) reported a skewed sex ratio in *Dinornis* (based primarily on morphometric data) ranging from 1 : 1 up to 1 : 6.1 between deposits, always in favour of females. Here, we used aDNA to unequivocally assign species and gender to a large sample of moa individuals from a restricted geographic region and a brief geologic time span. PV and BHV represent different depositional and taphonomic environments but were still “sampling” the same surrounding living moa assemblage, a situation that provided some significant comparative advantages and demonstrated how modern multidisciplinary dissections of fossil deposits can provide new insights into the biology of extinct fauna.

2.1.3 Regional setting

Pyramid Valley (42°58'22.0"S, 172°35'49.0"E) and Bell Hill Vineyard (42°58'19.36"S, 172°39'56.15"E) are 5.7 km apart, near Waikari, North Canterbury, South Island, New Zealand (Figures 2.1.1 and 2.1.2 and Digital Appendix 3). The PV deposit is characterised by exceptional bone preservation and high species richness. Most importantly, moa were preserved here – for the most part – as separate skeletons, which allowed discrimination between individuals. Moa fossils were discovered at PV in the late 1930s while the landowners were burying a horse. Extensive excavations between 1949 and 1973 involved ~7.5 % of the deposit's surface area, and the results were summarised by Holdaway and Worthy (1997).

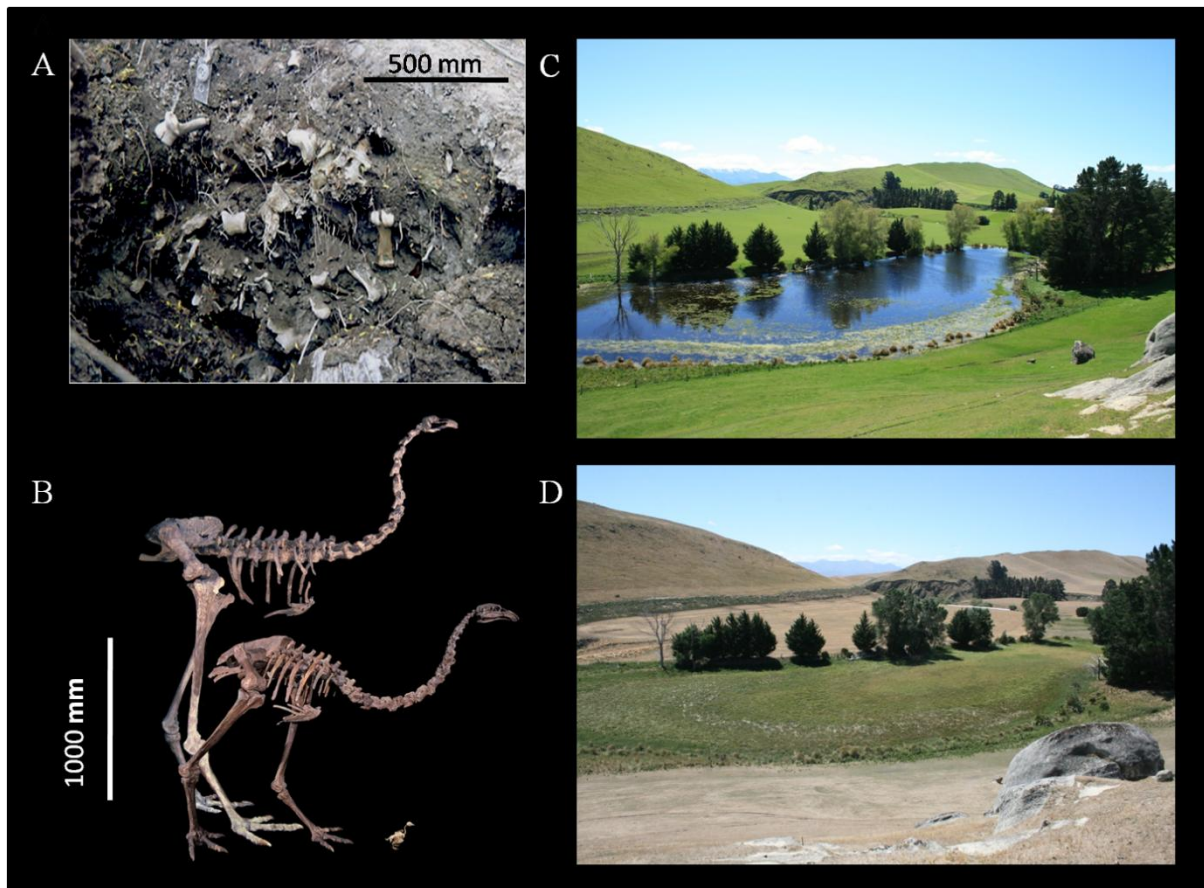


Figure 2.1.2: A: Fossil assemblage at BHV as exposed (and later reburied) during excavation in 2001. Bones of moa were superimposed in very high concentration. B: Extreme reverse sexual size dimorphism: female (left) up to 280 % heavier and 150 % taller than male (right) in *Dinornis robustus*, with skeleton of pigeon (*Columba livia*) for size comparison. C and D: PV, 2008. A shallow lake (C, October 2008, after two intense rain storms in July) is usually maintained at the site, but during prolonged drought the lake bed dries out (D, January 2008, southern summer). These seasonal and between-year differences in lake level were encountered by the local moa populations. In the late Holocene, prior to human arrival in New Zealand, the lake was surrounded by forest.

The PV deposit sampled the fauna inhabiting a seasonally dry forest on the lime stone hills surrounding a 1.5 ha shallow lake (Figure 2.1.2), and the fossils were preserved in the anoxic lake sediments (“gyttja”) (e.g., Burrows, 1989). The BHV fossil assemblage was discovered in 2001 during the excavation of a pond associated with a new vineyard. Much of the material had been displaced at the time of discovery but a significant sample of moa was excavated scientifically. The birds were trapped where a small stream crossed thixotropic Oligocene greensands amongst limestone hills (Turvey, Holdaway, 2005). Whereas individual moa were usually readily discernible in the PV sediments, BHV contained bones from numerous moa superposed in very high concentration (Figure 2.1.2). Fossils of more than 120 individuals were excavated from a 4 m² area of the deposit, which also contained fallen logs, an *in situ* stump, and root systems (RNH personal observation) suggesting

that the stream had flowed through a forest. Simple miring has been the preferred explanation for the presence of moa in the sites, and that probably applies for the BHV site. It is likely, however, that many moa at PV were killed by the giant extinct Haast's eagle (*Hieraaetus moorei*). At least 12 of the moa skeletons recovered from PV exhibit evidence (claw damage to the iliac plates of the pelvis) of predation by eagles (Worthy, Holdaway, 2002).

2.1.4 Materials and methods

2.1.4.1 Sampling

We sampled 268 bones (representing the same number of individuals) for aDNA analyses. This was all of the relevant PV and BHV material that could be located at Canterbury Museum, Christchurch (PV material, $n = 135$), the National Museum of New Zealand, Te Papa Tongarewa, Wellington (BHV material, $n = 116$), the American Museum of Natural History, New York (PV material, $n = 12$) and material temporarily held by RNH (BHV material, $n = 5$). All individuals are listed in Digital Appendix 4. To ensure that individuals were included just once in the analyses, only left tibiotarsi were sampled, except for a few PV individuals where only the right tibiotarsus was available. However, as individuals from PV were registered individually, with catalogue numbers allotted for entire skeletons and not individual bones, using a few right side elements from this site did not carry the risk of duplication. Also, one very small *D. robustus* chick from PV (AV 9063) was sampled from the tarsometatarsus. Sampling was conducted as in Allentoft et al. (2009), following strict aDNA protocols.

2.1.4.2 DNA extraction, species determination, and qPCR

DNA was extracted from 200 mg of bone powder (slightly less for some of the smallest juveniles) by incubation with rotation at 55°C for 48 h in 1.5 mL digestion buffer [20 mM Tris, pH 8.0, 1% Triton X-100, 10mM dithiothreitol (DTT), 1 mg/mL proteinase K and 0.47 M EDTA]. The supernatant was spun through 30,000 MWCO Vivaspin columns (Sartorius Stedim Biotech, Goettingen, Germany) and then combined with five volumes of PBI buffer (Qiagen, Valencia, CA) before the DNA was extracted using silica spin columns (Qiagen) and cleaned with AW1 and AW2 wash buffers (Qiagen). Finally the DNA was eluted in 50 µL of 10mM Tris buffer. The Mitochondrial Control region was amplified using a combination of primers (185F, 262F, 294R, 441R) as described previously (Bunce et al., 2003; Cooper et al., 2001). Diagnostic Control region single nucleotide polymorphisms (SNPs) were characterised from the amplified fragments and, when compared to ~300 moa GenBank

sequences (representing all nine species), we were able to identify and discriminate between the four moa taxa present at the PV and BHV sites (see Digital Appendix 5). Samples from individuals that yielded a different taxon from that listed in the museum catalogue were repeated in at least one additional PCR to confirm the result. Nomenclature of moa is still subject to changes as more genetic and other information becomes available: here we follow Bunce et al. (2009) in recognising a single species (*E. curtus*) in the genus *Euryapteryx*.

A large random subsample ($n = 228$) of the DNA extracts were screened with the 262F/441R primer combination in a quantitative PCR amplification setup using SYBR detection chemistry on a BioRad My-IQ thermocycler (Allentoft et al., 2009). The recorded cycle threshold (C_T) values provided a relative quantitative measure of the DNA preservation in each extract. In accordance with aDNA guidelines, DNA extractions and PCR setup were performed in separate, dedicated aDNA facilities at Murdoch University (Perth, Australia).

2.1.4.3 Molecular sexing

All birds display a ZZ/ZW sex chromosomal mechanism, with females being the heterogametic sex (e.g., Bloom, 1974; Clinton, Haines, 1999). We applied two independent molecular sexing protocols, amplifying moa specific W-linked sequences with lengths of 60 bp and 85 bp respectively. Primers (Moa1/Moa2 and KW1F/KW1R) were designed and loci amplified using PCR conditions described in Huynen et al. (2003) and Bunce et al. (2003) with the exception that Hi Fidelity Platinum Taq® polymerase was used instead of Gold Taq polymerase (both Invitrogen®) which increased the specificity of the reaction. Primer-dimer bands often interfered with the short W-linked PCR products on traditional gel-electrophoresis images, which made interpretations problematic. To avoid this, primers (Moa1 and KW1R) were 5'-labelled with fluorescent FAM dye (Integrated DNA Technologies, Coralville, IA) and the PCR products separated on an ABI 3730 genetic analyser by co-running a Genescan LIZ500 size standard (Applied Biosystems, Foster City, CA, U.S.A.). Negative extraction and amplification controls were always included. DNA fragments were scored manually with the aid of GENEMARKER® v 1.5 (Soft Genetics, State College, PA). The assigned sex of an individual was accepted only if it was confirmed by both PCR assays (Digital Appendix 6). Because of the risk that poorly preserved DNA would fail to amplify the W-linked sequences (and hence appear as males regardless of their true sex), all extracts were screened initially to confirm sufficient preservation of nuclear DNA. Primers amplifying a >100 bp moa-specific microsatellite (Moa_MS2) (Allentoft et al., 2009) were used for this purpose. Finally, extracts from all apparent males were subsequently repeated in PCR's by both sexing protocols to further minimize the risk of including

“false” males in the data. Individuals whose DNA yielded inconsistent results or did not amplify the Moa_MS2 locus were omitted from all analysis concerning sex ratios.

To allow for a comparison of sex ratios in juveniles and adults, the individual developmental stage was assessed for each sampled tibiotarsus according to the patterns of bone development and fusion described in Turvey and Holdaway (2005). The authors identified six different ontogenetic classes based on the state of development of the bone, but to avoid fragmenting the data beyond usage, we discriminated only between juveniles (stage I, II, III and IV) and adults (stage V and VI). The growth rate differed among moa higher taxa (Turvey et al., 2005) and the transition from juvenility to adulthood is gradual and obviously relies on the very definition of the terms. However, at stage V, the tibiotarsus of the birds had reached their full length (Turvey, Holdaway, 2005) so for the purposes of this study, we set the distinction at the easily separable features characterising stages IV and V. The ontogenetic age of a few bones with significant post-mortem damage or missing parts, could not be assessed (Digital Appendix 4). These individuals were excluded from the subsequent comparison of sex ratios in adults and juveniles.

2.1.5 Results

2.1.5.1 Species representation

Of 268 sampled bones, positive species identifications were obtained for 267. Only one individual (S 40302, listed in the museum catalogue as *E. crassus*) continually failed to amplify DNA. The species compositions are shown in Table 2.1.1 and Figure 2.1.3. Overall, *D. robustus*, *E. crassus* and *E. curtus* were equally common in the deposits (29 %, 30 % and 29 % respectively) whereas the largest emeid, *P. elephantopus*, contributed just 12 % of the individuals at both sites combined. The relative species abundances differed significantly between the two sites ($\chi^2_{df=3} = 215.9$, $P \ll 0.0001$). *D. robustus* and *E. crassus* dominated in the PV deposit whereas *E. curtus* outnumbered the other three taxa combined at BHV.

Only two individuals from PV had been misidentified on morphological criteria, but 21 individuals from BHV yielded genetic identifications at variance with morphological determinations. Of these, 14 genetically identified as *E. curtus* had been catalogued as *E. crassus*. Of the 28 sampled BHV tibiotarsi labelled simply “juvenile emeid”, 25 were identified genetically as *E. curtus*. The other three were *D. robustus* and hence dinornithids instead of emeids (Digital Appendix 4).

Table 2.1.1: Number of individuals (n) and relative abundance (overall and site-specific) of each moa taxon, based on aDNA analysis.

	Pyramid Valley		Bell Hill Vineyard		Overall	
	n	fraction	n	fraction	n	fraction
<i>D. robustus</i>	56	0.38	22	0.18	78	0.29
<i>E. crassus</i>	60	0.41	19	0.16	79	0.30
<i>E. curtus</i>	16	0.11	62	0.52	78	0.29
<i>P. elephantopus</i>	15	0.10	17	0.14	32	0.12
Total	147		120		267	

Table 2.1.2: Sex ratios (overall and site-specific) in each moa taxon as represented in the PV and BHV collections, based on aDNA analysis

	Pyramid Valley			Bell Hill Vineyard			Overall		
	♂	♀	♂ : ♀	♂	♀	♂ : ♀	♂	♀	♂ : ♀
<i>D. robustus</i>	2	47	1 : 23.5	6	11	1 : 1.8	8	58	1 : 7.3
<i>E. crassus</i>	7	53	1 : 7.6	2	13	1 : 6.5	9	66	1 : 7.3
<i>E. curtus</i>	0	13	-	16	25	1 : 1.6	16	38	1 : 2.4
<i>P. elephantopus</i>	0	15	-	4	13	1 : 3.3	4	28	1 : 7.0
Total	9	128	1 : 14.2	28	62	1 : 2.2	37	190	1 : 5.1

Sex ratios in bold differed significantly from equality (χ^2 , $\alpha = 0.05$)

2.1.5.2 Molecular sexing

Of the 267 individuals, from which mtDNA was successfully amplified, 227 (85 %) also contained sufficiently well-preserved nuclear DNA to warrant the use of molecular sexing techniques. The DNA extract from 40 individuals showed inconsistent results or failed to amplify the microsatellite and these were excluded from further analyses of the sex ratios. DNA from five individuals was extracted twice and amplification results from the independent extractions corresponded. Further, 17 of the moa individuals included here had been molecular sexed in previous studies in different laboratories (Bunce et al., 2003; Huynen et al., 2003) (Digital Appendix 4). Matching these results served as a successful independent replication, which is an important validation criterion in the field of ancient DNA (Cooper, Poinar, 2000). We tested for a relationship between cycle threshold (C_T) values (based on the 262F/441R mtDNA, qPCR amplifications) and the genetically-determined gender. If the absence of W-linked amplifications had indeed resulted from poor DNA preservation, this should be reflected in higher C_T values (i.e. a later DNA amplification) in males than among females in our tested subsample (228 randomly picked DNA extracts).

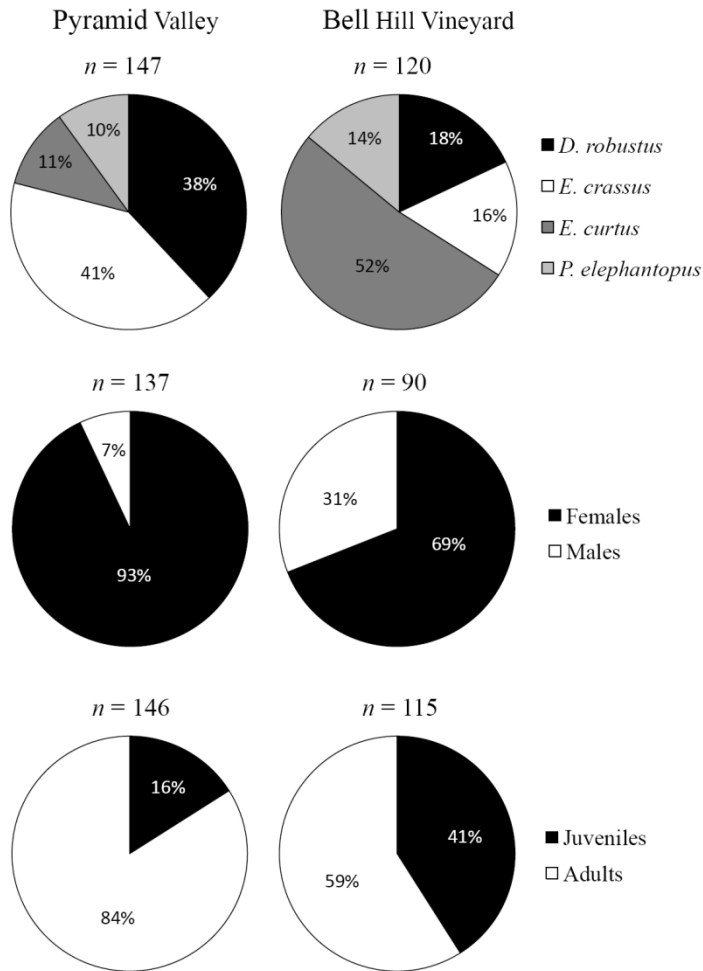


Figure 2.1.3: Pie charts showing extent of compositional differences in species abundances (upper panel), sex ratios (middle panel), and fraction of juveniles (lower panel), in the PV and BHV collections. 267 individuals were genetically identified to species: of these, 261 could be assigned to a maturity stage, as outlined in text; 227 individuals contained sufficiently preserved nuclear DNA to allow molecular sexing

A significant relationship between C_T value based on the mitochondrial 262F/441R primers and the ability to amplify nuclear DNA has already been established on a large proportion of our moa DNA extracts (Allentoft et al., 2009). The tested males did not display higher C_T values ($n = 33$, mean $C_T = 34.0$) compared to the females we tested ($n = 164$, mean $C_T = 34.2$) which indicated that the DNA preservation among detected males was just as good as among detected females (t -test: $t_{df=195} = 0.28$, $P = 0.78$). Individuals that could not be genetically sexed, however, displayed significantly (t -test: $t_{df=226} = 2.78$, $P = 0.006$) higher C_T values ($n = 31$, mean $C_T = 36.1$) than all the tested individuals that were successfully provided with a gender, male or female regardless ($n = 197$, mean $C_T = 34.1$). These associations strongly supported the results and displayed the robustness of the applied genetic sexing protocols.

Of the 227 individuals assigned a sex, 190 were females; and 37 were males (Table 2.1.2, Figure 2.1.3). Emphasizing the highly-skewed sex ratios, no males of *E. curtus* or *P. elephantopus* were detected in the sample from PV and only two males were identified among the 49 sexed *D. robustus* individuals from this site (Table 2.1.2). Overall, taxa and sites combined, the sex ratio ($\sigma : \phi$) was 1 :

5.1 and differed significantly from equality ($\chi^2_{df=1} = 93.7$, $P < 0.0001$). Sex ratios ranged from 1 : 1.6 within *E. crassus* at BHV to 1 : 23.5 within *D. robustus* at PV (Table 2.1.2), and in the far majority of test combinations among sites and species, the ratios differed significantly from equality in favour of females (Table 2.1.1). The sex ratios also deviated significantly ($\chi^2_{df=1} = 88.3$, $P < 0.0001$) between the two sites, with an overall 1 : 2.2 at BHV but 1 : 14.2 at PV (Table 2.1.2). A higher fraction of males at BHV compared to PV was evident within each taxon, and significantly so, except for *E. crassus* showing a somewhat similar ratio of 1 : 7.6 ratio at PV and a 1 : 6.5 ratio at BHV (Table 2.1.2). Although males were in clear minority at both sites, a much higher proportion of the moa at BHV were males.

Juveniles constituted a significantly ($\chi^2_{df=1} = 36.1$, $P < 0.0001$) higher fraction (41 %) of moa preserved at BHV than at PV (16 %) (Figure 2.1.3, Table 2.1.3). This resulted mainly from a very large number of juvenile *E. curtus* found at BHV (34 individuals) but also, 47 % of *D. robustus* at BHV were juveniles compared to just 18 % at PV (Table 2.1.3). The large number of *E. curtus* at BHV allowed for a meaningful sex ratio comparison between juveniles and adults within same species and site. Twenty juveniles and 21 adults of *E. curtus* from BHV were successfully sexed and the sex ratios proved significantly different ($\chi^2_{df=1} = 3.9$, $P = 0.049$), with a balanced 1 : 1 ratio among juveniles but a 1 : 2.5 ratio among adults (Table 2.1.3).

Table 2.1.3: Sex ratios in adults and juveniles of each moa taxon and for both sites as represented in the collections. Note the 1 : 1 sex ratio of juvenile *E. curtus* at BHV, in contrast to the 1 : 2.5 ratio in the adults.

Pyramid Valley	Juveniles					Adults					% juv	No age data	Total
	♂	♀	♂ : ♀	Not sexed	Total juv	♂	♀	♂ : ♀	Not sexed	Total ad			
<i>D. robustus</i>	0	9	-	1	10	2	38	1 : 19.0	6	46	18 %	0	56
<i>E. crassus</i>	0	10	-	0	10	6	43	1 : 7.2	0	49	17 %	1♂	60
<i>E. curtus</i>	0	3	-	1	4	0	10	-	2	12	25 %	0	16
<i>P. elephantopus</i>	0	0	-	-	0	0	15	-	0	15	0 %	0	15
Overall	0	22	-	2	24	8	106	1 : 16.0	8	122	16 %	1	147

Bell Hill Vineyard	Juveniles					Adults					% juv	No age data	Total
	♂	♀	♂ : ♀	Not sexed	Total juv	♂	♀	♂ : ♀	Not sexed	Total ad			
<i>D. robustus</i>	0	5	-	4	9	6	4	1.5 : 1	0	10	47 %	2♀/1?	22
<i>E. crassus</i>	0	1	-	1	2	2	11	1 : 5.5	3	16	11 %	1♀	19
<i>E. curtus</i>	10	10	1 : 1	14	34	6	15	1 : 2.5	7	28	55 %	0	62
<i>P. elephantopus</i>	0	2	-	0	2	4	10	1 : 2.5	0	14	14 %	1♀	17
Overall	10	18	1 : 1.8	19	47	18	40	1 : 2.2	10	68	41 %	5	120

Sex ratios in bold differed from equality

Molecular sexing of juveniles was generally less successful than for adults (30 % failure contra 9 % failure, respectively) (Table 2.1.3), which most probably reflected the smaller (less destructive) samples retrieved from the smallest bones. However, the C_T values of all the juveniles that did yield

satisfactory sexing data (and therefore used in further analyses), confirmed that the DNA was equally well-preserved in juvenile males and females, and in juveniles and adults (data not shown).

2.1.6 Discussion

2.1.6.1 Accuracy of morphological taxonomic identification

Two hundred and sixty seven of 268 bones yielded amplifiable DNA. This 99.6 % success ratio was substantially higher than the < 57.8 % (125 successful mtDNA amplifications from >230 moa bone samples) reported in Baker et al. (2005) or the 62.5 % (10 of 16 moa bone samples) in Huynen et al. (2008). We suspect that a combination of continuously-improved DNA isolation protocols (see Rohland, Hofreiter, 2007) and exceptional preservation of DNA in the geologically young bones from PV and BHV resulted in DNA yields better than described in previous moa aDNA studies.

Accurate species identifications are essential for interpreting the history and taphonomy of any fossil deposit. The DNA-based protocol utilized here provided accurate identifications of nearly all moa individuals excavated from two large Quaternary deposits in New Zealand. One convenient by product of this approach was that it allowed us to assess the accuracy of the taxonomic labelling (based on morphological criteria) in the sampled collections. The overall observed rate of misidentification (8.7 %) in these collections proved lower than the 33.6 % proposed by Baker et al. (2005) and much closer to the 4 % suggested by Scofield et al. (2005). Only two PV individuals yielded different identifications, but 21 BHV bones had been taxonomically misplaced on morphological data (Digital Appendix 4). Many of these were, however, specimens lacking clear diagnostic features because they were damaged or from small juveniles, or both. Of 13 misidentified BHV bones that could be genetically sexed, nine were determined to be males of a different taxon than that proposed on morphological data (Digital Appendix 4). This clearly reflects how sexual dimorphism has contributed to the (smaller) males being wrongly assigned to a smaller taxon. The inclusion of genetically-determined juveniles in the BHV sample significantly altered the relative species composition and effectively doubled the proportion of identified *E. curtus* at that site. These results suggest that previous analyses of relative species abundance from sites with many juveniles could be affected by systematic errors and that material used for detailed studies of moa biology would benefit from being identified genetically, if the DNA preservation allows it and especially if the sample includes juveniles.

2.1.6.2 Overall species abundances

The analyses showed that four moa species occupied the area around the two sites in the late Holocene environment of North Canterbury, which accords with earlier non-DNA assessments of the local prehistoric avian fauna (Holdaway, Worthy, 1997; Worthy, Holdaway, 1996). Based on pooled data from both sites *D. robustus*, *E. crassus*, and *E. curtus* were equally represented with ~ 30 % each, whereas *P. elephantopus* contributed only 12 % of the individuals (Table 2.1.1). Such low representation at both sites suggests that *P. elephantopus* was the least common moa in the area. This species – the largest of the emeids – is commoner in early Holocene fossil assemblages from other North Canterbury sites. This incongruence has been interpreted (Worthy, Holdaway, 1996) as reflecting an association with more open shrubland habitats, becoming restricted as forests spread after the most recent (Otiran) glaciation in New Zealand. *E. curtus*, on the other hand, is not well represented in older North Canterbury sites, and the same authors proposed that this taxon required a more closed forest habitat and increased in abundance in North Canterbury as forest composition changed through the late Holocene. These scenarios are plausible, but the significant compositional differences demonstrated here between PV and BHV, despite their proximity and contemporaneity, highlight some of the problems associated with such interpretations and suggests that more subtle features of the environment were controlling moa distribution, abundance, and deposition.

2.1.6.3 Niche separation, biased deposition or unrepresentative collections?

The relative abundances of moa taxa in the two adjacent sites proved very different; an unexpected extent of geographical distribution heterogeneity for such large and highly mobile animals. A range of factors may have caused the apparent pattern of *D. robustus* and *E. crassus* being deposited in the sediments of PV whereas *E. curtus* dominated at BHV <6 km away. First and most easily assessed, the assemblages may have been biased by trade and exchange of material so that the material available for sampling is no longer representative of the primary collections (Hunter, Donovan, 2005). However, all material from BHV was accounted for at the recent (2001) and only excavation and practically all material was available for sampling, apart from a few elements held in a private collection. Holdaway and Worthy (1997) conducted a thorough assessment of the PV collection at Canterbury Museum and concluded that the majority of the material collected during the many excavations spanning 40 years, was still at the museum. The largest number of individuals held elsewhere, 12 skeletons or partial skeletons at the American Museum of Natural History (New York) were all included in the present study. Although a few catalogued bones in Canterbury Museum could not be located during our sampling programme, it seems highly unlikely that exchange and curatorial biases have had any significant influence on our results.

The relative abundances within the collections could also be biased by site-specific or species-specific differences in the likelihood of deposition, and clearly, the very different taphonomies at the two sites constituted different ways of sampling the same moa fauna. Potential depositional and taphonomic biases providing a skewed reflection of the surrounding prehistoric ecosystem are well known challenges in palaeological reconstructions (e.g., Badgley, 1986; Kidwell, 2001; Staff et al., 1986). Behavioural aspects of the fauna can also bias deposition, as likely illustrated in, for example, the Rancho La Brea deposits in California, where large carnivores outnumber herbivores by nine to one (e.g., Carbone et al., 2009; Spencer et al., 2003). Despite Atkinson and Milleners' (1991) hypothesis that heavier moa taxa such as *D. robustus* and *P. elephantopus* would have been over-represented in deposits such as PV and BHV, there seems to be no mass-dependent deposition bias because the largest and heaviest (*D. robustus*) and smallest and lightest (*E. crassus*) moa in the area were recovered in equal proportions at PV and BHV (Table 2.1.1, Figure 2.1.3). The PV (Holdaway, Worthy, 1997) and BHV fossil assemblages (RNH unpubl. data) also included small (< 1 kg) non-moa taxa which further supports the contention of a largely indiscriminate deposition.

We argue, therefore, that it is unlikely that there was sufficient interspecific bias in the mode of deposition of individuals (from a theoretical homogenous assemblage of living moa) at either or both sites that would account for the observed differences in their relative abundances at PV and BHV. It seems more reasonable that the 'availability' of each taxon for deposition at the two sites would have differed, for example as a result of differences in habitat preferences. Although most of North Canterbury was clothed in woody vegetation in the late Holocene, a mosaic of forest and shrubland would have offered a heterogeneous array of edge and subcanopy habitats (e.g. Batcheler, 1989; Burrows, 1989). The significant morphological variation between moa species in traits such as body size and beak shape combined with differing distribution ranges certainly indicate some level of ecological separation (Atkinson, Greenwood, 1989; Batcheler, 1989; Cooper et al., 1993; Worthy, Holdaway, 2002). Perhaps reflecting such niche partitioning, the smallest (*E. crassus*) and largest (*D. robustus*) of the local moa appear to have been co-abundant at PV whereas both were less frequent at BHV, where the intermediate-sized *E. curtus* dominated the overall moa fauna. Although no convincing evidence based on the contents of coprolites and gizzards (Burrows, 1989; Wood et al., 2008) has been presented for niche partitioning, stable isotopic data do suggest such variation between taxa in moa (Worthy, Holdaway, 2002). A detailed discussion of the possible nature and composition of the suggested difference in microhabitats is beyond the scope of this paper, but habitat around a small stream in the forest (BHV) probably differed somewhat from that around the 1.5-ha lake at PV and could have prompted different behavioural responses (attraction / avoidance) among the four species of moa in the area.

2.1.6.4 Sex ratios

Sixty-five percent of 200 published estimates of adult sex ratios in birds differed significantly from equality (Donald, 2007). Although there may be a considerable publication bias (Donald, 2007), unbalanced sex ratios are clearly not uncommon amongst birds. Most bird species with unbalanced sex ratios display, however, an excess of males (Donald, 2007), whereas our data showed an excess of females. In the following, we distinguish between primary, secondary, and tertiary sex ratios, being the ratios of male to female in eggs, chicks, and independent non-juvenile individuals, respectively, as defined by Mayr (1939). There is a substantial literature on the adaptive advantages that could explain variation in offspring sex allocation (e.g. Hardy, 2002; Sheldon, 1998; West et al., 2002), but we have presented evidence that the sex ratio was not skewed in juvenile *E. curtus* and although sufficient data were available for a meaningful comparison between adults and juveniles of that taxon alone, the result is unambiguous and significant. It clearly demonstrates an absence of imbalance in the primary or secondary sex ratios of *E. curtus* individuals but a 1 : 2.5 ratio among adults from the same site. Perhaps compromising this observation as a general feature in moa, no males were detected among the 27 sexed juveniles which were not *E. curtus* (from the two sites combined) (Table 2.1.3). This possible discrepancy between sex ratios in juveniles of *E. curtus* and juveniles of the other taxa may, in part, reflect the relative immaturity (maturity class I) of many sexed juveniles of *E. curtus* from BHV. If the factor(s) causing a tertiary sex skew did not operate on the earliest life stages, all these very young *E. curtus* individuals have not been affected yet, whereas the pooled sample of juveniles from the other taxa included all the immature bone developmental classes (I to IV) (data not shown).

If the museum collections are representative of the original collections it leaves at least two plausible explanations for an apparently female-biased adult sex ratio: moa developed female-biased tertiary sex ratio as a result of higher male mortality; or disproportionately more females than males, from populations with balanced sex ratios, were deposited in the sites. A male-biased tertiary sex ratio is common in birds, most likely because adult females in general suffer higher mortality as a result of higher metabolic demands and greater risk of predation during incubation (e.g., Donald, 2007; Thomson et al., 1998; Trivers, 1972). Territorial males also exclude females from the most favourable habitats which can result in differential survival (Benkman, 1997; Marra, Holmes, 2001; Pechacek, 2006). Consequently, in birds displaying reversed sex roles and unusual mating systems, adult females have in some instances been observed to outnumber adult males (Donald, 2007). All modern flightless ratites (*Struthio*, *Dromaius*, *Casuarius*, *Rhea*, *Pterocnemia*, *Apteryx*), and at least some of the volant tinamous, exhibit greater or lesser degrees of reversal in their sex roles, with parental care often wholly or partly undertaken by the males (Handford, Mares, 1985). The southern cassowary (*Casuarius casuarius*), perhaps the best living behavioural analogue of moa, exhibits significant (up to 50 %) reverse sexual dimorphism (RSD), is usually solitary and territorial, and females dominate males and may mate with more than one male each year (Crome, 1975). Moa, as large ratites with a

high degree of RSD, could have exhibited reversed sex roles, with adult males experiencing higher mortality rates resulting from the costs associated with parental care, including prolonged fasting during incubation, and exclusion from the most productive habitat by territorial females. One indication for greater longevity of females is the much lower adult: juvenile ratio in males than in females of *E. curtus* (Table 2.1.3). Despite reversed sex roles in modern ratites, the literature is equivocal on there being a general excess of females, with reports of male-biased, female-biased, and balanced sex ratios (Coddington, Cockburn, 1995; Donald, 2007; Moore, 2007).

Although a hypothesis of higher adult male mortality would explain the overall excess of females documented in this and previous studies, it fails to explain why sex ratios differed significantly between the two sites. Clearly, one or other, or both deposits, cannot be recording the true sex ratios, because they were both “sampling” from the same moa assemblage in space and time. With females of *D. robustus* being up to 280 % heavier than males (Bunce et al., 2003), a mass-dependant bias at PV may seem preferable to explain the extreme sex skew observed in this deposit. However, the overwhelming female excess is tempered by the observation that the largest (*D. robustus*) and smallest (*E. crassus*) moa in the area were found in almost equal proportions at this site (38 % and 41 %, respectively), so mass-related deposition is unlikely to have caused the observed patterns. Instead, behavioural differences between the sexes could have caused a sex-related pattern of deposition. Holdaway & Worthy (1997) suggested that a different representation of sexes in the Pyramid Valley deposit may be attributed to a difference in male and female behaviour around the water source. Waterholes and their surrounding habitat often represent valuable resources to large territorial vertebrates and for example, male springbok (*Antidorcas marsupialis*) defend territories beside waterholes, which results in extremely skewed mating success (Ritter, Bednekoff, 1995). If large female moa were strongly territorial, only few males would have had access to the water in that area, which would limit the risk for males to be incorporated in the deposits. Strong female territorial behaviour by an extant ratite – the cassowary – has been captured on film (BBC, 2009) and Crome (1975) reported that a dominant cassowary female only needed to be “stretching and staring silently” at the males for them to promptly leave her vicinity. If PV was included in the home range of a dominant female moa, it might be expected that fewer males and juveniles would be represented in the deposit. A single territorial male defending a harem of many females near the lake could also produce the observed skew in the deposit, but we consider that the degree of RSD precluded male dominance in moa.

In addition to a territorial-related deposition bias, paternal egg incubation in moa, as suggested in Birchard and Deeming (2009), could also lower the probability of males being incorporated in the deposit. Worthy and Holdaway (1996) suggested that most of the moa at PV died during the driest period over summer (Figure 2.1.2) when birds might have been tempted out on to a substantially dry lake bed, where a thin crust covered the “gyttja”, to find water. Gizzard contents of birds from PV

contained seeds and fruits (Burrows et al., 1981), perhaps supporting the contention of a summer or autumn deposition (Holdaway, Worthy, 1997; Turvey, Holdaway, 2005). If adult males were incubating elsewhere during summer, and then occupied in taking care of the hatchlings within the cover of the forest, they were less likely to be incorporated in the deposit. Likewise, the low numbers of juveniles for all taxa at PV may have resulted from a preference for heavier cover during maturation, as known for *Dromaius* where the male takes the young into woodlands where there is a lower risk of predation by raptors (Marchant, Higgins, 1990). Such seasonal, reproduction-related differences in activity and habitat could contribute to both the observed patterns the overall female excess, and the differences between the sites, if indeed the small BHV site was in dense forest as suggested by the plant material – including in situ stumps – preserved in the site (RNH pers. obs).

The sex ratios at the two sites have been discussed previously but without the accuracy provided by genetic identifications. Exemplifying this, Holdaway & Worthy (1997) proposed that both sexes of *E. crassus* were found in equal proportions at PV, whereas we demonstrate the presence of just seven males compared to 53 females. Bunce et al. (2003) estimated a balanced 1 : 1 sex ratio within *D. robustus* from BHV, but at that site, we identified six males and 11 females of this taxon. The true extent of the suggested tertiary gender-imbalance in the four sympatric moa populations remains unknown, but the overall 1 to 2.2 female-biased sex ratio revealed at BHV is not unique among birds (Donald, 2007) and certainly seems more realistic than the exceptional 1 to 14.2 ratio of the pooled sample from PV. However, our results also emphasize the caution that should be exercised in assessing the biology of moa in general terms without discriminating between taxa. The four species examined occupied the same area, but must have differed significantly in several biological aspects to produce the fossil assemblages discussed here. To fully understand the real sex ratios in the living populations of each taxon, accurate genetic assessments of taxon and sex are needed from a range of fossil sites, including cave deposits and pit fall traps (Worthy, Holdaway, 1993; Worthy, Holdaway, 1994; Worthy, Holdaway, 1995; Worthy, Mildenhall, 1989) that perhaps sampled the surrounding moa populations without some of the potential biases at lake-, bog, and swamp sites. In addition, juveniles must be included in the analyses. Turvey and Holdaway (2005) concluded that age class distribution and sex ratios in the sample from PV do not represent equal sampling of an autochthonous population of *Dinornis*. Our genetic analysis suggests that that holds true for all four moa taxa at PV. Our results also supported Turvey and Holdaway's (2005) contention that the BHV assemblage is more likely to reflect the composition of the surrounding *Dinornis* population, because our assessment showed that there generally was a much higher fraction of juveniles and males at BHV than at PV.

2.1.7 Conclusions

This study was based on the largest ancient DNA inventory ever obtained from a geographically and temporally constrained megafaunal fossil assemblage and it has again demonstrated that aDNA is a powerful source of information when biomolecule preservation permits. The species compositions and sex ratios presented here could not have been quantified without aDNA technology.

The analyses revealed significant compositional differences between the deposits in relative species abundance, sex ratios, and frequency of juveniles, despite the two sites being contemporaneous and just 5.7 km apart, over easy terrain. We argue that moa behaviour and the different local microhabitats are responsible for the observed patterns. It is clear, however, that all four sympatric moa taxa in the late Holocene of North Canterbury exhibited a considerable excess of adult females. The data for *E. curtus* indicate that this reflects, at least partly, a tertiary female-biased sex ratio. We hypothesize that the skewed sex ratio was driven by higher rates of mortality among adult or sub-adult males, or both, resulting from the reversed sex roles that characterise extant large ratites. The extreme excess of females observed at PV could reflect this site's being a valuable water and food resource, defended assiduously by large dominant females. Paternal incubation and parental care, as in extant ratites, would also have isolated the adult males from the open lake bed and further reduced the probability of their deposition.

In summary, the genetic evidence presented here has provided new information on the biology of moa populations in the prehuman New Zealand ecosystem. Once these genetic data are analysed in relation to stable isotope signatures and a radiocarbon chronology of the birds (ongoing studies), unparalleled insight into the palaeoecosystem will emerge. Moreover, recent advances in the identification of highly variable nuclear microsatellites in moa (Allentoft et al., 2009) have the potential to allow an assessment of local-scale population dynamics of New Zealand's lost megafauna in much higher resolution than attempted previously. Finally, this study clearly demonstrates the caution required when interpreting single fossil deposits in isolation as they may present a biased picture of the living prehistoric fauna. The different taphonomies of the two sites provided opportunities for comparative analysis and insights at levels that would not have been available for sites in isolation.

2.1.8 Acknowledgements

We gratefully acknowledge the Museum of New Zealand, Te Papa Tongarewa (A.J.D.Tennyson), Canterbury Museum, The American Museum of Natural History, and Sherwyn Veldhuizen and Marcel Giesen for allowing sampling of moa in their collections. We thank the Hodgen family for their long-term and continuing commitment to the Pyramid Valley site, its contents and their support

of our research projects. We thank Malene Møhl, Emma McLay, Charlotte Oskam, James Haile, Ross Barnett and Tom Gilbert for help during the sampling. We would also like to thank Trevor Worthy, Alan Cooper and the Natural History Museum (Oxford) for assistance with production of the moa skeleton image. Financial support was provided by the Marsden Fund of the Royal Society of New Zealand (Contract 06-PAL-001-EEB, Palaecol Research Ltd), and the Australian Research Council.

2.1.9 References

- Allentoft ME, Schuster S, Holdaway RN, et al. (2009) Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *Biotechniques* **46**, 195-200.
- Atkinson IAE, Greenwood RM (1989) Relationships between Moas and Plants. *New Zealand Journal of Ecology* **12**, 67-96.
- Atkinson IAE, Millener PR (1991) An ornithological glimpse into New Zealand's pre-human past. *Acta XX Congressus Internationalis Ornithologici* **1**, 129-192.
- Badgley C (1986) Counting individuals in mammalian fossil assemblages from fluvial environments. *Palaios* **1**, 328-338.
- Baker AJ, Huynen LJ, Haddrath O, Millar CD, Lambert DM (2005) Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: The giant moas of New Zealand. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8257-8262.
- Barnes I, Shapiro B, Lister A, et al. (2007) Genetic Structure and Extinction of the Woolly Mammoth, *Mammuthus primigenius*. **17**, 1072-1075.
- Barnett R, Shapiro B, Barnes I, et al. (2009) Phylogeography of lions (*Panthera leo* ssp.) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. *Molecular Ecology* **18**, 1668-1677.
- Batcheler CL (1989) Moa browsing and vegetation formations, with particular reference to deciduous and poisonous plants *New Zealand Journal of Ecology* **12**, 57-65.
- BBC (2009) Cassowaries. In: *BBC Natural World*.
- Benkman CW (1997) Feeding behavior, flock-size dynamics, and variation in sexual selection in crossbills. *Auk* **114**, 163-178.
- Birchard GF, Deeming DC (2009) Avian eggshell thickness: scaling and maximum body mass in birds. *Journal of Zoology* **279**, 95-101.
- Blau M, Deevey ES, Gross MS (1953) Yale Natural Radiocarbon Measurements, I. Pyramid Valley, New Zealand and Its Problems. *Science* **118**, 1-6.
- Bloom SE (1974) Current knowledge about the avian W-chromosome. *Bioscience* **24**, 340-344.
- Bunce M, Worthy TH, Ford T, et al. (2003) Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425**, 172-175.
- Bunce M, Worthy TH, Phillips MJ, et al. (2009) The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20646-20651.

- Burrows CJ (1989) Moa browsing - evidence from the Pyramid Valley mire. *New Zealand Journal of Ecology* **12**, 51-56.
- Burrows CJ, McCulloch B, Trotter MM (1981) The diet of moas based on gizzard contents samples from Pyramid Valley, North Canterbury and Scaifes Lagoon, Lake Wanaka, Otago, New Zealand. *Records of the Canterbury Museum* **9**, 309-336.
- Carbone C, Maddox T, Funston PJ, et al. (2009) Parallels between playbacks and Pleistocene tar seeps suggest sociality in an extinct sabretooth cat, *Smilodon*. *Biology Letters* **5**, 81-85.
- Clinton M, Haines LC (1999) An overview of factors influencing sex determination and gonadal development in birds. *Cellular and Molecular Life Sciences (CMLS)* **55**, 876-886.
- Coddington CL, Cockburn A (1995) The mating system of free-living emus. *Australian Journal of Zoology* **43**, 365-372.
- Cooper A, Atkinson IAE, Lee WG, Worthy TH (1993) Evolution of the Moa and Their Effect on the New-Zealand Flora. *Trends in Ecology & Evolution* **8**, 433-437.
- Cooper A, Lalueza-Fox C, Anderson S, et al. (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704-707.
- Cooper A, Mourer-Chauvire C, Chambers GK, et al. (1992) Independent Origins of New Zealand Moas and Kiwis. *Proceedings of the National Academy of Sciences* **89**, 8741-8744.
- Cooper A, Poinar HN (2000) Ancient DNA: Do it right or not at all. *Science* **289**, 1139-1139.
- Crome FHJ (1975) Some observations on the biology of the cassowary in northern Queensland. *Emu* **76**, 8-14.
- Debruyne R, Chu G, King CE, et al. (2008) Out of America: Ancient DNA Evidence for a New World Origin of Late Quaternary Woolly Mammoths. *Current Biology* **18**, 1320-1326.
- Donald PF (2007) Adult sex ratios in wild bird populations. *Ibis* **149**, 671-692.
- Gilbert MTP, Drautz DI, Lesk AM, et al. (2008) Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 8327-8332.
- Gregg D (1972) Holocene stratigraphy and moas at Pyramid Valley, North Canterbury, New Zealand. *Records of the Canterbury Museum* **9**, 151-158.
- Haddrath O, Baker AJ (2001) Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proceedings of the Royal Society of London Series B-Biological Sciences* **268**, 939-945.
- Handford P, Mares MA (1985) The mating systems of ratites and tinamous - an evolutionary perspective. *Biological Journal of the Linnean Society* **25**, 77-104.
- Hardy ICW (1997) Possible factors influencing vertebrate sex ratios: An introductory overview, 217-241.
- Hardy ICW (2002) Sex ratios: Concepts and research methods. *Sex ratios: Concepts and research methods*, i-xiii, 1-424.
- Hofreiter M, Rabeder G, Jaenicke-Després V, et al. (2004) Evidence for Reproductive Isolation between Cave Bear Populations. *Current Biology* **14**, 40-43.
- Holdaway RN, Jacomb C (2000) Rapid extinction of the moas (Aves : Dinorinithiformes): Model, test, and implications. *Science* **287**, 2250-2254.
- Holdaway RN, Worthy TH (1997) A reappraisal of the late quaternary fossil vertebrates of Pyramid Valley Swamp, North Canterbury, New Zealand. *New Zealand Journal of Zoology* **24**, 69-121.
- Hunter AW, Donovan SK (2005) Field sampling bias, museum collections and completeness of the fossil record. *Lethaia* **38**, 305-314.
- Huynen L, Lissone I, Sawyer S, Lambert D (2008) Genetic identification of moa remains recovered from Tiniroto, Gisborne. *Journal of the Royal Society of New Zealand* **38**, 231-235.

- Huynen L, Millar CD, Scofield RP, Lambert DM (2003) Nuclear DNA sequences detect species limits in ancient moa. *Nature* **425**, 175-178.
- Kidwell SM (2001) Preservation of species abundance in marine death assemblages. *Science* **294**, 1091-1094.
- Kulp JL, Tryon LE, Eckelman WR, Snell WA (1952) Lamont Natural Radiocarbon Measurements, II. *Science* **116**, 409-414.
- Marchant S, Higgins PJ (1990) *Handbook of Australian, New Zealand and Antarctic Birds: Ratites to ducks* Oxford University Press, Melbourne.
- Marra PP, Holmes RT (2001) Consequences of dominance-mediated habitat segregation in American Redstarts during the nonbreeding season. *Auk* **118**, 92-104.
- Mayr E (1939) The sex ratio in wild birds. *American Naturalist* **73**, 156-179.
- Moore LA (2007) Population ecology of the southern cassowary *Casuarius casuarius johnsonii*, Mission Beach north Queensland. *Journal of Ornithology* **148**, 357-366.
- Orlando L, Bonjean D, Bocherens H, et al. (2002) Ancient DNA and the Population Genetics of Cave Bears (*Ursus spelaeus*) Through Space and Time. *Mol Biol Evol* **19**, 1920-1933.
- Pechacek P (2006) Foraging behavior of Eurasian Three-toed Woodpeckers (*Picoides tridactylus alpinus*) in relation to sex and season in Germany. *Auk* **123**, 235-246.
- Rawlence NJ, Wood JR, Armstrong KN, Cooper A (2009) DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa. *Proceedings of the Royal Society B: Biological Sciences* **276**, 3395-3402.
- Ritter RC, Bednekoff PA (1995) Dry season water, female movements and male territoriality in springbok: Preliminary evidence of waterhole-directed sexual selection. *African Journal of Ecology* **33**, 395-404.
- Rohland N, Hofreiter M (2007) Comparison and optimization of ancient DNA extraction. *Biotechniques* **42**, 343-352.
- Scofield P, Worthy T, Holdaway RN, et al. (2005) Recent claims for more moa and huge errors in museum identifications - cutting through the spin. *Journal of Museums Aotearoa* **30**, 29-31.
- Shapiro B, Drummond AJ, Rambaut A, et al. (2004) Rise and fall of the Beringian steppe bison. *Science* **306**, 1561-1565.
- Sheldon BC (1998) Recent studies of avian sex ratios. *Heredity* **80**, 397-402.
- Spencer LM, Van Valkenburgh B, Harris JM (2003) Taphonomic analysis of large mammals recovered from the Pleistocene Rancho La Brea tar seeps. *Paleobiology* **29**, 561-575.
- Staff GM, Stanton RJJ, Powell EN, Cummins H (1986) Time-averaging, taphonomy, and their impact on paleocommunity reconstruction: Death assemblages in Texas bays. *Geological Society of America Bulletin* **97**, 428-443.
- Thomson DL, Monaghan P, Furness RW (1998) The demands of incubation and avian clutch size. *Biological Reviews* **73**, 293-304.
- Trivers RL (1972) Parental investment and sexual selection In: *Sexual Selection and the Descent of Man: The Darwinian Pivot* (ed. Campbell B), pp. 136-179. Aldine Pub. Co., Chicago.
- Turvey ST, Green OR, Holdaway RN (2005) Cortical growth marks reveal extended juvenile development in New Zealand moa. *Nature* **435**, 940-943.
- Turvey ST, Holdaway RN (2005) Postnatal ontogeny, population structure, and extinction of the giant moa *Dinornis*. *Journal of Morphology* **265**, 70-86.
- West SA, Reece SE, Sheldon BC (2002) Sex ratios. *Heredity* **88**, 117-124.
- Wood JR, Rawlence NJ, Rogers GM, et al. (2008) Coprolite deposits reveal the diet and ecology of the extinct New Zealand megaherbivore moa (Aves, Dinornithiformes). *Quaternary Science Reviews* **27**, 2593-2602.

- Worthy TH (1987) Sexual dimorphism and temporal variation in the North Island moa species *Euryapteryx curtus* (Owen) and *Pachyornis mappini* (Archy). *National Museum of New Zealand Records* **3**, 59-70.
- Worthy TH, Holdaway RN (1993) Quaternary fossil faunas from caves in the Punakaiki area, West Coast, South Island, New Zealand. *Journal of the Royal Society of New Zealand* **23**, 147-254.
- Worthy TH, Holdaway RN (1994) Quaternary fossil faunas from caves in Takaka Valley and on Takaka Hill, Northwest Nelson, South Island, New Zealand. *Journal of the Royal Society of New Zealand* **24**, 297-391.
- Worthy TH, Holdaway RN (1995) Quaternary Fossil Faunas from Caves on Mt-Cookson, North-Canterbury, South-Island, New-Zealand. *Journal of the Royal Society of New Zealand* **25**, 333-370.
- Worthy TH, Holdaway RN (1996) Quaternary fossil faunas, overlapping taphonomies, and palaeofaunal reconstruction in north Canterbury, South Island, New Zealand. *Journal of the Royal Society of New Zealand* **26**, 275-361.
- Worthy TH, Holdaway RN (2002) *The lost world of the Moa* Canterbury University Press, Christchurch.
- Worthy TH, Mildenhall DC (1989) A Late Otiran Holocene Paleoenvironment Reconstruction Based on Cave Excavations in Northwest Nelson, New-Zealand. *New Zealand Journal of Geology and Geophysics* **32**, 243-253.

Degradation of ancient DNA in bone – insights from three megafaunal sites in New Zealand

2.2.1 Abstract

In January 2008 an assemblage of large ancient bones was unearthed in a field near Waikari, South Island, New Zealand. We describe this new fossil site; Rosslea, and provide an inventory of the excavated material. Ancient DNA was isolated from 14 bones and a piece of eggshell. Genetic species identifications matched those based on morphology, confirming three species of New Zealand's extinct megaherbivore – the moa (Aves: Dinornithiformes) – and the extinct South Island Adzebill (*Aptornis defossor*). Eight bones from Rosslea were ^{14}C AMS dated and median calibrated dates ranging from 7839 to 1482 years BP, showed a Holocene fossil accumulation. The presence of three fossil-rich sites (Rosslea, Pyramid Valley and Bell Hill Vineyard) in close proximity enabled us to conduct a comprehensive comparative study of ancient DNA preservation over a significant geologic age range. A quantitative PCR (qPCR) experiment involving 158 radiocarbon-dated moa bones provided clear empirical evidence that ancient DNA degradation follows an exponential decay process as previously hypothesised. However, we observed large variability in DNA preservation between bones and sites, which, probably explains why previous studies have failed to demonstrate a temporal trend. We estimated a DNA half-life of 521 years for the targeted 250 bp fragment of mtDNA, but the decay rate varied from site to site. The Rosslea bones displayed a relatively poor DNA preservation, and with a shorter estimated molecular half-life (389 years). Similar climatic conditions prevail at the three sites, which are separated by < 5 km, so the results suggest that factors such as sediment chemistry have been affecting long-term DNA preservation. This study demonstrates the utility in conducting molecular characterisations in parallel with more classical paleontological approaches.

2.2.2 Introduction

Since its inception in the mid 1980s (Higuchi et al., 1984; Paabo, 1985a; Paabo, 1985b) the field of ancient DNA (aDNA) has progressed from a scientific novelty, mastered by a few visionary pioneers, to represent an established field, contributing to research in a range of paleontological, geological and archaeological contexts. It is no longer controversial that the genetic information preserved in ancient substrates such as bone, mummified flesh, hair, feathers, eggshell, coprolites, and sediments can be extracted and add to our understanding of the past. Several recent contributions have emphasised the potential in collaborations between molecular biologists, archaeologists, and palaeontologists

(Allentoft et al., 2010; Barnett et al., 2009; Bramanti et al., 2009; Bunce et al., 2009; Debruyne et al., 2008; Gilbert et al., 2008; Hebsgaard et al., 2009; Oskam et al., 2010). Therefore, when a large assemblage of well-preserved, non-mineralised fossil bones was recently unearthed at Rosslea, near Waikari in North Canterbury on the South Island of New Zealand (see Figure 2.2.1), samples were obtained and aDNA technology and radiocarbon dating applied as integrated components in a description of the Rosslea site.

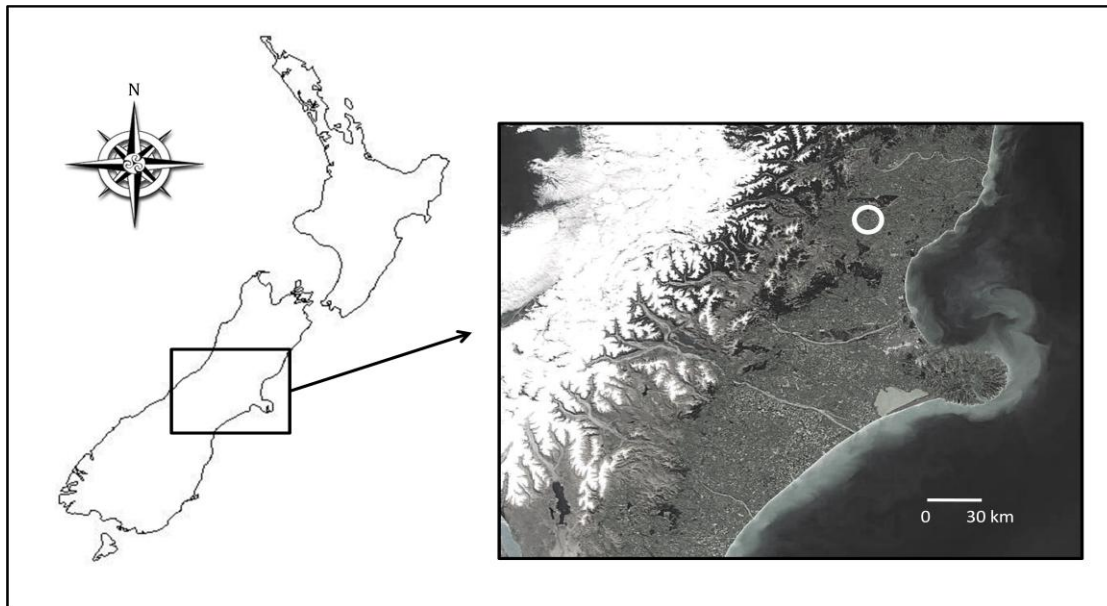


Figure 2.2.1: Map of New Zealand with central South Island highlighted. White circle frames the Waikari area in North Canterbury where Rosslea, Pyramid Valley and Bell Hill Vineyard are located. See Figure 1.3.1 or Digital Appendix 3 for relative locations of the three study sites. Satellite image was downloaded and modified (with permission) from the European Space Agency (<http://www.esa.int>).

North Canterbury has a rich Holocene fossil record, dominated by four extinct moa species (Aves: Dinornithiformes) (Worthy, Holdaway, 1996), and the landowner responsible for the discovery recognized immediately that the bones were moa remains. However, identification of moa remains to species level can sometimes be challenging. Nine species are presently recognized (Bunce et al., 2009), but sexual dimorphism (Bunce et al., 2003; Huynen et al., 2003) and intra-species size variation between different geographic regions at different times (Worthy, 1987; Worthy, Holdaway, 1996), complicates identification. Trained palaeontologists rarely misidentify well-preserved bones of adult moa, but the results presented in Allentoft et al. (2010) demonstrated that a molecular approach represents the best (if not only) way of providing a secure species ID for juveniles and damaged moa bones, lacking clear diagnostic features. The study also showed that a disproportionately large fraction

of bones from the (smaller) males had been morphologically misidentified as smaller taxa. These results emphasized how aDNA analyses can increase the quantity and quality of the information available from a fossil site.

Using quantitative PCR (qPCR), molecular analyses of the Rosslea assemblage was extended beyond species identification, to include a characterization of DNA preservation in the bones. DNA degrades with time in dead tissue and a range of mechanisms (summarised in Lindahl, 1993) have been suggested to explain this post-mortem instability. In this study, we assessed mainly *non-bypassable* DNA damage (Schwarz et al., 2009), which prevents the polymerase enzyme in the PCR reaction to continue beyond the damaged site. This type of damage therefore limits the number of suitable template DNA molecules, recognized as a later offset in the monitored qPCR amplification reaction. Although interstrand DNA or DNA-protein crosslinks might account for most of the non-bypassable DNA damage in some instances (Hansen et al., 2006; Poinar et al., 1998), it is generally accepted that the rate of molecular decay (i.e. DNA half-life) is normally determined by hydrolytic damage causing baseless sites followed by strand breakage. This process of random DNA strand fragmentation is believed to follow a 1st order kinetics model, and be responsible for an exponentially decreasing number of suitable DNA template molecules with time (e.g., Deagle et al., 2006; Paabo, Wilson, 1991). This assumption has been applied occasionally in attempts to predict long term survival of DNA (Marota et al., 2002; Paabo, Wilson, 1991; Pruvost et al., 2007; Smith et al., 2003).

Despite theoretical expectations, many studies have struggled to demonstrate empirically a simple correlation between age and DNA quality in historical or ancient samples (Colson, Bailey, 1997; Dobberstein et al., 2008; Gilbert et al., 2003; Hagelberg et al., 1991; Haynes et al., 2002; Hoss et al., 1996; Paabo, 1989; Poinar et al., 1996; Schwarz et al., 2009). This failure is likely to reflect the need for a large, homogenous, and well-dated sample to overcome a considerable noise, generated by physical, chemical, and biological variables such as deposition temperature, microbial activity, pH, ionic strength, presence of oxygen and chemically-free water, which can all affect the rate of DNA decay (Lindahl, Andersson, 1972; Lindahl, Nyberg, 1972; Lorenz, Wackernagel, 1994; Smith et al., 2003). To investigate the levels and rates of molecular degradation, we generated a large dataset comparing DNA preservation in the Rosslea material with moa bones from two adjacent North Canterbury fossil sites, Pyramid Valley (PV) and Bell Hill Vineyard (BHV). Quantitative PCR data from 158 radiocarbon-dated moa bones (Holdaway et al. In prep), allowed us to assess time-dependant molecular degradation with a much larger dataset than has been available before. Further, the proximity of the three sites (<5 km apart) effectively eliminated any confounding effects of climatic variables on DNA degradation between samples, and provided a level of homogeneity lacking in previous investigations.

Here, we first provide an overview of the material recovered from Rosslea with morphological species identifications supported by analyses of ancient DNA sequences. We advocate that DNA preservation should be regarded as an important taphonomic character as it highlights the potential for further molecular work at a given site, as well as provides a measure of the integrity of the material recovered. We report comparisons between the molecular preservation at the three adjacent and largely contemporary North Canterbury sites. We also assess whether a proposed signal of time-dependant exponential DNA degradation can be confirmed, and whether meaningful molecular half-life values can be estimated. Finally, we discuss within-site and between-site differences in DNA preservation.

2.2.3 Materials and methods

2.2.3.1 Discovery and description of the site

The Rosslea fossil assemblage was discovered by the landowner in January 2008, during excavation with a large mechanical digger of a water storage dam in a known spring location (42°57'53.83''S, 172°39'22.39''E). A significant excavation was underway at the renowned Pyramid Valley site (4.8 km away), and the researchers here were made aware of the discovery. The fossils had been removed from 9 m² but most had been found in a 1.7 m deep, rectangular (c. 3.5 x 1.5 m) pit in the south-western corner of the excavated area (Figure 2.2.2). The site was on the northern side of a slight rise in an otherwise horizontal surface. Geologically, the site is at the junction of Holocene river gravels with silt and fan deposits, and a small exposure of the Oligocene Waima Formation, manifested as an undifferentiated limestone. The pit itself showed significant lateral differentiation with both silt and peaty soils. The bones were found primarily in a matrix of hemic peaty soil. As the surrounding substrate was primarily sandy, it is believed that the presence of peat near the bone bed was a result of the spring's increased productivity. With an extensive excavation being carried out at Pyramid Valley, it was not possible to mobilize the required work force and expertise for a detailed excavation at Rosslea. Instead, a series of bones (Table 2.2.1) was donated to Canterbury Museum for further examination, cataloguing and sampling. The remainder of the recovered material remains at Rosslea, stored under cool, humid conditions. No controlled excavations has been undertaken at Rosslea and the pit now serves as part of a water storage pond.



Figure 2.2.2: The Rosslea site as it appeared shortly after the bones were discovered in January 2008. *Right:* The deep rectangular pit where the majority of bones were found. *Left:* Many fossils were recovered from the spoil heap in the background.

2.2.3.2 Sampling

In addition to the traditional physical description and morphology-based identification of each fossil, samples were obtained for molecular and biochemical analyses. There is clear evidence that post mortem DNA degradation intensifies after bones are removed from their deposition environment (Pruvost et al., 2007), so all samples were obtained within 14 days of the initial exposure. A total of 14 bones (representing the same number of individuals) was sampled. One femur, 12 left tibiotarsi and 1 tarsometatarsus from a South Island adzebill (*Aptornis defossor*) were obtained (Table 2.2.1). A piece of fossil eggshell was also sampled (Oskam et al., 2010). Samples were removed from each bone, using a hollow diamond-coated drill bit (Diamond Drill Bit & Tool Co., Omaha, USA; model DT-516; OD, 12.9 mm; ID, 10.9 mm) driven by a hand-held electric drill. To minimize potential effects of within-bone heterogeneity in DNA content (e.g., Kaiser et al., 2008), we always sampled the cortical bone from the central shaft. Samples were stored at 12°C before further processed. A second core was removed from the head of each bone (Figure 2.2.3) and stored at 12°C for future reference. The analyzed plug of bone was prepared with a Dremel (Racine, WI, USA) cutting disk (part no. 409) as follows: A top slice was removed and split in half (Figure 2.2.3), one half for stable isotopic analyses (ongoing study) and the other for AMS ^{14}C dating. The remaining cortical bone mass was ground into powder with a Dremel drill bit (Part no. 114), to increase the surface area for efficient aDNA extraction. To prevent cross-sample contamination, sampling tools and the sampling enclosure were cleaned thoroughly between samples, as described in Allentoft et al. (2009).

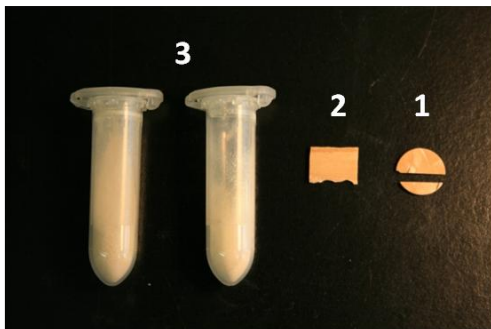


Figure 2.2.3: Sampling moa fossils.

Cores ($d = 1.1$ cm) were removed from the bone. The top disk of each core was separated from the rest and split for radiocarbon dating and stable isotope analyses (1). A median longitudinal section of the core (2), was removed for individual maturity assessments (ongoing study). Because of poor preservation state, this could not be achieved from Rosslea fossils. Remaining cortical bone mass was ground into powder (3) for aDNA extraction

2.2.3.3 ^{14}C AMS dating

Bone gelatin was extracted from samples of bone by Isolytix Ltd, Dunedin, New Zealand, using standard protocols. Gelatin sample splits were sent for stable isotopic analysis (Environmental Isotopes Ltd, Sydney, Australia) and AMS radiocarbon analysis. The gelatins were processed according to the protocol in Holdaway & Beavan (1999) at GNS Science, Lower Hutt, New Zealand. Radiocarbon ages were measured at GNS Science. Ages are reported as conventional radiocarbon ages, and as median calibrated ages and 95% probability distributions generated by the OxCal v4.1 (Oxford Radiocarbon Accelerator Unit) program. Sample quality was assessed using the C: N molar mass ratio, and stable isotopic ratios ($\delta^{15}\text{N}$, $\delta^{13}\text{C}$) in comparison with archive data.

2.2.3.4 Ancient DNA extraction, PCR and preservation

DNA extractions from 200 mg aliquots of bone powder and PCR amplifications were performed at dedicated aDNA facilities (Murdoch University, Perth, Australia) as described previously (Allentoft et al., 2010; Allentoft et al., 2009). To ensure secure species identification, we used a combination of primers (185F, 294R, 262F, 441R) targeting mitochondrial Control region sequences (Bunce et al., 2003; Cooper et al., 2001). PCR products were sequenced in both directions, “blasted” against the ~ 300 moa Control region sequences available on GenBank. Using this approach, the species identity of each bone could be confirmed with high fidelity. Sequences displaying rare mutations were repeated and confirmed in at least one additional PCR before accepted. For PCR products with ambiguous

base-calls, the amplified DNA was cloned to investigate the problem using a TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA), and sequenced using vector specific M13 primers according to the manufacturer's instructions. DNA from the Rosslea eggshells had already been extracted and identified by Oskam et al. (2010).

DNA amplifications of the 262F/441R fragment (250 base pairs including primers) were performed in a quantitative PCR (qPCR) setup, using SYBR-green detection chemistry on a BioRad My-IQ thermocycler. This involved all the 158 DNA extracts included this study. The recorded cycle threshold (C_T) values allowed us to assess the relative DNA preservation in each extract. The quantitative accuracy in this setup has been validated before (Allentoft et al., 2009; Oskam et al., 2010), but to assess the level of inconsistency between qPCR reactions, 22 randomly chosen extracts were repeated.

The number of DNA copies in a PCR reaction follows an exponential increase per cycle, hence the relative amounts of DNA in each extract can be calculated. For example, if DNA extract "A" produces a C_T value three cycles higher than extract "B" then the former has $2^3 = 8$ times less DNA. Using the 262/441 qPCR assay we compared DNA preservation in Rosslea fossils with that of material from PV and BHV. Using a suite of 158 radiocarbon ages, we assessed and compared DNA preservation in fossils from different times.

Further, we tested for the ability to amplify nuclear DNA. Each cell contains only one nuclear genome but many mitochondria, so "traditionally" aDNA research has targeted mitochondrial DNA fragments. We used primers for a nuclear moa-specific microsatellite locus (Moa_MS2) which has previously been shown to amplify well in North Canterbury moa fossils (Allentoft et al., 2009).

Preliminary data suggested that one of the four North Canterbury moa species (*Pachyornis elephantopus*) amplified later in the qPCR reaction than the other species, regardless of age (data not shown). This could be a primer binding issue or DNA being harder to extract from the very dense bones of this taxon. To avoid potential bias in the analyses, data for this species (not present in the Rosslea sample) were excluded. In total, we assessed DNA preservation in 158 radiocarbon-dated moa fossils, including the 8 dated bones from Rosslea. Museum accession numbers and other information linked to the BHV and PV material are found elsewhere (Allentoft et al., 2010) while the AMS ^{14}C data will be presented in Holdaway et al. (In prep).

2.2.4 Results

2.2.4.1 Morphological description

All the catalogued and sampled fossils recovered from Rosslea are listed in Table 2.2.1. Bones from at least 17 moa individuals were assessed: 14 left-tibiotarsi were excavated (12 of which were sampled), a separate *Dinornis* right femur (CM AV41188, also sampled), and a left femur from a moa chick (CM AV41198, not sampled) (Table 2.2.1). Lastly, associated leg bones from an adult female *Dinornis* remain at the site and were not sampled. Three species of moa were present, with material of *Euryapteryx curtus* being most abundant, and slightly fewer *Emeus crassus* elements. The site contained at least three individuals of *Dinornis robustus*, of which two were sampled. The two sampled individuals of *Dinornis* differed in size, with the early (7839 BP; CM AV41188) individual being considerably larger than the younger unaccessioned specimen (4403 BP, designated “R8” here). Other taxa present included a single, partially-articulated *Aptornis defossor*, and *Gallirallus australis*. Fragments of egg shell c. 1 mm thick, and with the distinctive slit shaped pores diagnostic of moa (Oskam et al., 2010) were also retrieved.

The bones from Rosslea were stained deep brown, consistent with preservation in peat (Figure 2.2.4). The cortical bone appeared porous, with a tendency for cores to split during sample drilling. Their texture was very different from the solid and cohesive cortical bone of moa material from PV and BHV.



Figure 2.2.4: Colour difference between fossils from the different sites. Proximal ends of two left tibiotarsi of eastern moa (*Emeus crassus*) are shown. Upper bone (R13) represents a Rosslea element, and lower bone (CM AV8340) was recovered from PV.

Table 2.2.1: Overview of the material from Rosslea that was recovered and assessed. Fourteen bones (13 moa, one *Aptornis defossor*) and one piece of eggshell were sampled and processed for aDNA analyses. Eight moa bones were ^{14}C AMS dated. The median calibrated radiocarbon ages are given in years before present (i.e. before 1950) along with the 95% probability distribution limits. GenBank accession numbers are provided for all DNA sequences. Material that remained at Rosslea has not been assessed for maturity level. “Amb” refers to DNA sequences with ambiguous base calls that could not be resolved through cloning. These ambiguities, however, were not present at taxonomically informative sites, and did not affect the ability to genetically assign the sampled bones to species level. “Short” refers to DNA extracts that did not produce a clean sequence across the entire targeted mtDNA CR fragment (185F/441R) and was therefore trimmed and submitted to GenBank as a shorter sequence. Again, this did not affect the ability to determine the taxon. Abbreviations: R fem, right femur; L tbt, left tibiotarsus; Tmt, tarso-metatarsus; Fib, fibula; Vert, vertebrae; Ad, adult; Juv, juvenile; Imm, immature; Emeid, indeterminate member of the moa family Emeidae, CM = Canterbury Museum; R = Rosslea; * elements used to estimate MNI; ** presented in Oskam et al. (2010).

Catalogue no.	Taxon (morphology)	Confirmed by aDNA	Element	Maturity	Sample no.	Calib. ^{14}C age	95% CI	Molecular sexing	GenBank no.
At Rosslea, R8	<i>D. robustus</i>	Yes	L tbt*		R8/MA235	4403 BP	4416-4164	Male	JF357912
CM AV41188	<i>D. robustus</i>	Yes, amb.	R fem*	Ad	R1/MA228	7839 BP	7938-7722	Failed	JF357919
CM AV41189	<i>E. curtus</i>	Yes	L tbt*	Juv	R2/MA229	5812 BP	5899-5662	Failed	JF357916
CM AV41190	<i>E. curtus</i>	Yes, short	L tbt*	Imm	R3/MA230			Failed	JF357923
CM AV41191	<i>E. curtus</i>	Failed	L tbt*	Ad	R4/MA231			Failed	
CM AV41193	<i>E. curtus</i>	Yes	L tbt*	Ad	R6/MA233	1482 BP	2110-930	Female	JF357917
At Rosslea, R9	<i>E. crassus</i>	Yes	L tbt*		R9/MA236	5703 BP	5885-5605	Failed	JF357913
At Rosslea, R10	<i>E. curtus</i>	Yes	L tbt*		R10/MA237	5514 BP	5583-5329	Female	JF357918
At Rosslea, R11	<i>E. crassus</i>	Yes, short	L tbt*		R11/MA238	5692 BP	5856-5605	Failed	JF357920
At Rosslea, R12	<i>E. crassus</i>	Yes, amb.	L tbt*		R12/MA239			Failed	JF357921
At Rosslea, R13	<i>E. crassus</i>	Yes	L tbt*		R13/MA240			Failed	JF357914
At Rosslea, R14	<i>E. crassus</i>	Yes, amb.	L tbt*		R14/MA241			Failed	JF357922
At Rosslea, R15	<i>E. crassus</i>	Yes	L tbt*		R15/MA242	5791 BP	5987-5599	Failed	JF357915
CM AV41198	Emeid		L fem*	Pullus					
CM AV41199	Emeid		L tbt*	Juv					
CM AV41203	<i>E. curtus</i>		L tbt*	Ad					
CM AV41192	<i>E. curtus</i>	Sampled but not processed	R tbt	Ad	R5/MA232				
CM AV41195	<i>E. curtus</i>		L&R tmt	Juv					
CM AV41196	<i>E. curtus</i>		L&R fem	Juv					
CM AV41197	<i>E. curtus</i>		R tbt	Juv					
CM AV41200	<i>E. curtus</i>		L&R tmt	Ad					
CM AV41201	<i>E. curtus</i>		L&R fem	Ad					
CM AV41202	<i>E. curtus</i>		R tbt	Imm					
CM AV41204	<i>E. curtus</i>		4 fib	Ad					
CM AV41205	<i>E. curtus</i>		R tmt	Ad					
CM AV41206	Emeid		10 vert	Ad					
CM AV41207	Emeid		14 phalanges	Ad					
CM AV41208	Moa sp.		8 sternal ribs	Ad					
CM AV41209	<i>D. robustus</i>		2 vert	Ad					
CM AV41210	<i>D. robustus</i>		R fib	Ad					
Murdoch Uni	<i>E. crassus</i>	yes	eggshell		RL_E1				GU799594 **
CM AV41194	<i>Aptornis defossor</i>	yes	R tmt	Ad	R7/MA234				JF357911
CM AV41211	<i>Gallirallus australis</i>		R hum	Ad					

2.2.4.2 ^{14}C AMS ages

Conventional radiocarbon ages (CRA) on moa bone gelatin from Rosslea ranged from 7001 ± 45 BP to 1540 ± 270 BP (Table 2.2.1). Their median calibrated ages corresponded to a depositional period from 7839 BP to 1482 BP deposition timeframe (95% probability limits: 7938 BP to 930 BP) (Table

2.2.1). Although the 8 dated moa bones from Rosslea included an interval >6300 years, five fell within a 300-year span (c. 5500 BP - 5800 BP).

2.2.4.3 Genetic species determination

We successfully amplified DNA from 13 of the 14 bones (93%) and one piece of eggshell, sampled for genetic taxon determination. In all 13 bones, the genetic identification matched the morphological assignment, confirming the presence of three moa species and a South Island adzebill (Table 2.2.1). Only one moa fossil (CM AV 41191, identified morphologically as, *Euryapteryx curtus*) continually failed to amplify DNA. Of the 13 successful amplifications, eight samples provided clean DNA sequences across the entire 380 bp fragment and in several independent PCRs (GenBank accession numbers in Table 2.2.1). Two samples could not amplify the entire 380 bp region but the shorter amplification could still be used for species identification (Table 2.2.1). Another three samples yielded slightly inconsistent DNA sequences, with rare mutations that could not be confirmed in subsequent PCR reactions. This clearly indicated post mortem DNA damage, causing miscoding lesions (e.g., Gilbert et al., 2007; Gilbert et al., 2003). Further PCRs and cloning did not resolve these issues. Despite providing secure species identifications, the exact DNA sequences of these three remained uncertain and were therefore submitted to GenBank with ambiguities.

2.2.4.4 qPCR results and nuclear DNA preservation

C_T values were recorded for the eight radiocarbon-dated Rosslea bones. These spanned 16.4 cycles and ranged from $C_T = 30.5$ in AV 41193 to $C_T = 46.9$ in AV 41188 being the youngest and oldest fossil respectively (Table 2.2.1).

The 22 repeated qPCRs (material from all three sites) showed an average deviation in C_T value of 1.07, but this was influenced by two samples yielding rather large deviations (5.23 and 3.92 cycle shift respectively). 17 of the 22 repeats deviated from the first PCR with less than one cycle, confirming the general consistency in the setup. There was, as expected, a trend towards higher variability in results from poorly preserved samples with low copy numbers (i.e. C_T values >40).

The mean C_T value for Rosslea was 39.6 and proved significantly higher than corresponding values for BHV with a mean C_T of 31.7 (t -test, unequal variances, $t_{df=8} = 4.0$, $P = 0.004$) and PV with an mean C_T of 34.1 (t -test, unequal variances, $t_{df=7} = 2.9$, $P = 0.022$) (Table 2.2.2). Furthermore, also BHV and PV fossils displayed a significant difference in mean C_T values. This was tested overall (t -test, equal variances, $t_{df=148} = 4.5$, $P < 0.001$) using data of all AMS dated fossils from the entire

deposition period of the two sites in combination, but it was also tested within their period of temporal overlap, framed by a shorter deposition span in BHV (Table 2.2.2). Eliminating age as contributing factor, PV fossils still displayed significantly higher C_T values than BHV material (t -test, unequal variances, $t_{df=72} = 3.8$, $P < 0.001$).

The average C_T values were converted to DNA copy-numbers, relative to the average value at BHV, set to 100%. PV fossils contained on average 19.6% (i.e. 80.4% less) of the template molecules present in BHV fossils and the mean relative copy number in a Rosslea bone was just 0.3% (Table 2.2.2). A simple “knockout” experiment, adding moa DNA from Rosslea and BHV to the same PCR reaction, rejected PCR inhibition as the cause of the high C_T -values at Rosslea. Three PCRs with combined extracts (1 μ l BHV DNA extract + 1 μ l Rosslea DNA extract) amplified just as efficiently as a PCR with a pure BHV extract.

For nuclear DNA amplifications, most PV and BHV bones were successful (87% and 94% respectively), whereas only three Rosslea bones (AV 41193, R8 and R10) amplified the nuclear *Moa_MS2* locus (Table 2.2.2). With a confirmation of nuclear DNA preservation, these three individuals were then genetically sexed following the strict, conservative procedure outlined in Allentoft et al. (2010), using moa female-specific primers (Bunce et al., 2003; Huynen et al., 2003). The two *Euryapteryx* individuals (AV41193, R10) were females and the *Dinornis* (R8) was a male (Table 2.2.1), which is consistent with its small size.

2.2.4.5 DNA preservation through time

To test for an association between DNA preservation and age, the relative proportion of template molecules was calculated in relation to the best sample (*D. robustus*, S39955, $C_T = 24.8$) set to 100%. First, we tested whether the observed degradation through time would be described better by linear or exponential decay, by regressing age against observed values and log-transformed values respectively. Molecular preservation as a simple linear function of age was not supported ($R^2 = 0.001$, ANOVA $_{df=157}$: $F = 0.27$, $P = 0.60$) whereas log-transformed values showed a highly significant regression ($R^2 = 0.386$, ANOVA $_{df=157}$: $F = 98.27$, $P < 0.001$) (Figure 2.2.5). This result suggested strongly that molecular degradation in these bones followed a 1st-order kinetic pattern (i.e. exponential decay). The degradation pattern was then investigated in the three sites separately.

Table 2.2.2: Average measures of DNA preservation across three sites and 158 radiocarbon dated moa fossils. “Mean C_T ” refers to the average value across all qPCR amplifications and “DNA copies” is the number of DNA templates, relative to the average value for BHV fossils (set to 100%). $t_{1/2}$ is the estimated molecular half-life for the here targeted fragment. “MS2 amp” denotes the proportion of fossils from which the nuclear microsatellite locus Moa_MS2 (Allentoft et al 2009) could be amplified. *Lower table:* To minimise the influence of geological age, the same measures were compared across the ~2000 year period (851 BP to 2876 BP) where bones accumulated concurrently at BHV and PV.

Entire timeframe						
	n	Mean age (BP)	Mean C_T	DNA copies	$t_{1/2}$	MS2 amp.
BHV	47	1923 (851-2876)	31.7	100%	-	87%
PV	103	1970 (602-4136)	34.1	19.6%	506 yrs	91%
Rosslea	8	5280 (1482-7839)	39.6	0.3%	389 yrs	23%
Overall	158	2123 (602-7839)	33.7	-	521 yrs	85%

BHV vs PV in overlapping timeframe				
n	Mean age (BP)	Mean C_T	DNA copies	MS2 amp.
47	1923 (851-2876)	31.7	100%	87%
70	1679 (915-2797)	33.7	25.0%	94%

After log-transformation, PV ($R^2 = 0.359$, ANOVA $df = 102$: $F = 56.53$, $P < 0.001$) and Rosslea ($R^2 = 0.749$, ANOVA $df = 7$: $F = 17.91$, $P = 0.005$) both displayed a highly significant regression, whereas the same pattern was not statistically supported for BHV fossils ($R^2 = 0.058$, Anova $df = 46$, $F = 2.79$, $P = 0.10$). Having identified exponential decay, we estimated the average molecular half-life of the amplified Control region fragment, using the exponential decay relation: $N_t = N_0 * e^{-kt}$, (N_t and N_0 being the quantity at time t and 0 respectively, and k the decay constant). The average half-life ($t_{1/2} = (\ln 2)/k$) for PV was 506 years, 389 years for Rosslea, and overall for all three sites combined: 521 years (Table 2.2.2). Since we could not detect an exponential decay process for the BHV material, no half-life was estimated for this site.

2.2.5 Discussion

2.2.5.1 Age, species-composition and taphonomy

The Rosslea fossil assemblage is a typical late Holocene site, containing bones of New Zealand’s megafauna from the last few millennia preceding extinction. Polynesians colonized New Zealand in

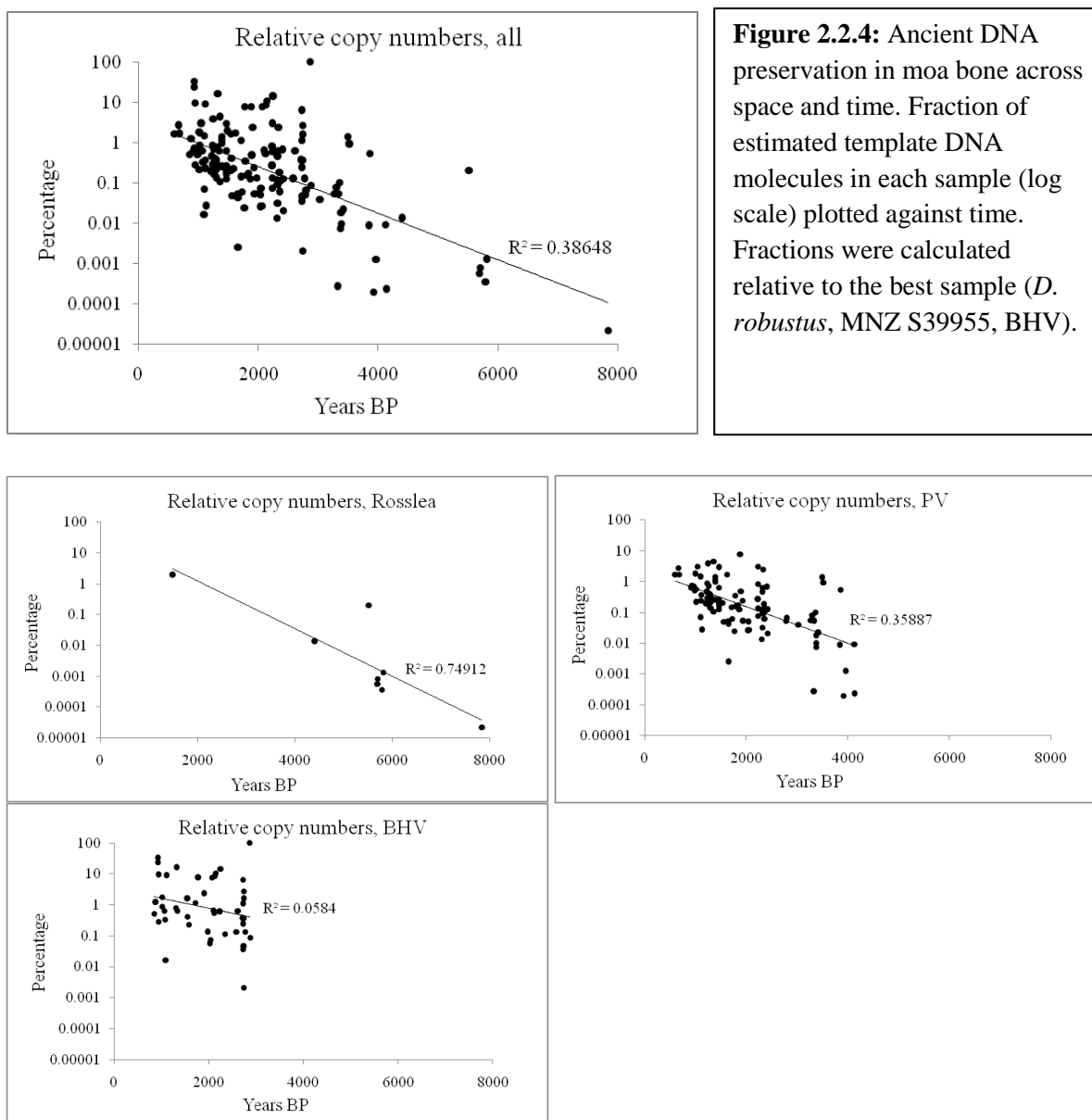
the late 13th century (Higham et al., 1999) and moa extinction followed within 150 years after that (Holdaway, Jacomb, 2000). Although we have only assessed a fraction of the material in the deposit, the faunal composition and dating results are consistent with previous evidence that the principle moa species in the area at this time were *Dinornis robustus*, *Emeus crassus* and *Euryapteryx curtus* (Allentoft et al., 2010; Holdaway, Worthy, 1997; Wood et al., 2008; Worthy, Holdaway, 1996; Worthy, Holdaway, 2002). Fossils of *Pachyornis elephantopus* are frequently encountered in North Canterbury sites, but its absence from the Rosslea material assessed here, supports the hypothesis that this was the rarest of the “eastern four” sympatric species in this region for that time (Allentoft et al., 2010). Because no proper scientific excavation was done at Rosslea, only limited information on the deposition depth, bone orientation, and other features was available, so it was impossible to thoroughly assess the taphonomy. However, the apparent dense accumulation of bones, combined with a relative scarcity of *Dinornis* and high frequency of *Euryapteryx* (Table 2.2.1) suggest that the site “sampled” the local moa populations by mirroring, similar to the spring-fed stream at the Bell Hill Vineyard site, 3.2 km away (Allentoft et al., 2010; Wood et al., 2008).

In contrast, the moa fauna of the Pyramid Valley lake deposit is heavily dominated by adult *Dinornis*, and characterized by an extreme sex bias (only 2 males among 48 genetically sexed *Dinornis* individuals), possibly because the productive edge vegetation around the lake served as a territorial domain for large females (Allentoft et al., 2010). Despite limited success in genetically sexing the Rosslea bones, we were able to show that at least one of the three excavated *Dinornis* (one sampled, at Canterbury museum; one sampled, at site; one large unsampled, at site) was male, so an extreme excess of females is unlikely at Rosslea.

Although a longer depositional time span was demonstrated at Rosslea, than at PV and BHV (Table 2.2.2), five of the eight dated individuals died within just 300 years (5500 BP to 5800 BP), suggesting that deposition was most intense in that relatively short period.

2.2.5.2 Genetic species identification and DNA preservation through time

Despite being several thousand years old, 13 of 14 Rosslea bone samples and a piece of eggshell generated DNA sequences of sufficient quality for secure species identification. The high success rate confirms the fine preservation state of ancient biomolecules encountered in late Pleistocene and Holocene fossils from New Zealand. Not all sequences were equally clean, however, with some showing ambiguous bases. In addition, several single-base mutations in otherwise clean sequences were identified as unique when queried against all GenBank moa sequences, but proved to be DNA damage artefacts when they were not confirmed in subsequent cloning and PCR reactions.



The observed damage patterns were consistent with Type 2 miscoding lesions (C→T/ G→A) which are encountered frequently in aDNA research (Gilbert et al., 2007; Gilbert et al., 2003; Hansen et al., 2001; Olivieri et al.), resulting mainly from hydrolytic damage causing deamination of cytosine (e.g., Lindahl, 1993).

The qPCR results clearly demonstrated that the number of template molecules in the fossils had decreased exponentially, with highly significant regressions in the log-transformed data. In this study, we could not distinguish between the different types of non-bypassable damage (i.e. crosslinks or depurination), but the observed exponential decay pattern corresponded well with a hypothesised first order kinetics model of random fragmentation. Spontaneous hydrolysis leads to depurination (loss of purines) of the DNA strand, which greatly destabilizes the DNA molecule and consequently leads to

strand cleavage by a β -elimination reaction (e.g., Lindahl, 1993). A considerable effort has been expended to investigate such time-dependent aDNA degradation, usually unsuccessfully (see Introduction), but Hansen et al. (2006) demonstrated an increase in DNA lesion frequency with increasing age of ancient permafrozen sediments, from which the DNA was extracted. Bar et al. (1988) showed that the length of the DNA fragments that could be amplified from human tissues (blood, muscle, spleen) decreased exponentially in the 20 days after a person's death. This accords, although on a very different timescale, with the patterns presented here.

Deagle et al. (2006) used a qPCR approach to assess the different proportions of DNA fragment lengths in samples of sea lion faeces and showed that the number of DNA fragments followed an exponentially declining curve as a function of their length. The same pattern was confirmed in an exhaustive assessment of non-bypassable DNA degradation in six permafrozen mammoth bones (Schwarz et al., 2009). Although these studies measured DNA preservation within single samples at only one point in time, rather than comparing the number of fixed-length template molecules in samples of different ages, they dealt with the same effect: if strand breaks appear randomly by depurination, large PCR-amplifiable DNA fragments will be disproportionately rare at any given time after cell death, because they contain many consecutive phosphodiester bonds with an equal probability of breaking.

2.2.5.3 Within-site variation in DNA preservation

Although the overall pattern of exponential DNA degradation was unambiguous, only 36 % of the variation could be accounted for by age alone (Figure 2.2.5). Separately, age accounted for 31 % at PV and 75 % at Rosslea. A higher explanatory value at Rosslea could be a simple sampling effect, given that only 8 data points are available from this site, but rather, it is an effect of the longer temporal window, allowing age to dominate the relationship. At BHV, geologic age was a very poor predictor of DNA quality, explaining just 2 % of the variation. Several factors may have obscured the expected link. For example, the moa recovered from BHV, were deposited over c. 2000 years, which is a relatively short period over which to examine the association.

Small juvenile moa often display poorer DNA preservation than adults (Allentoft et al., 2010), probably because their less dense bone provides limited protection to endogenous bio-molecules. Hence, the more ontogenetically heterogeneous fossil assemblage at BHV, with many juveniles and many (smaller) males than at PV, which is dominated by adult females (Allentoft et al., 2010), is likely to be more variable in its preservation. However, removing juveniles from the analyses did not improve the regression (data not shown).

Local taphonomic characteristics can also cause within-site differences in preservation (Hagelberg et al., 1991; Haynes et al., 2002; Smith et al., 2003). For example, if sediment deposition suddenly stopped, leaving the youngest fossils more exposed at the surface of the deposit for a prolonged period, an age-dependent degradation signal could be obscured. Moreover, it is not unlikely that other forms of non-bypassable DNA damage (such as cross-links) are present, and that degradation by these processes deviate from the 1st order kinetics expected by depurination alone. Radiocarbon uncertainties add to the noise in the system, but the errors are an order of magnitude less than the age ranges involved (see Digital Appendix 2).

Lastly, we note that some heterogeneity between DNA extractions in terms of yields and presence of inhibitors and stochastic variation in the subsequent PCR reactions is practically inevitable (as demonstrated from the 22 PCR repeats), even when strict, consistent, and standardised molecular procedures are followed.

2.2.5.4 Between-site variation in DNA preservation

An average Rosslea sample contained just 0.3% (i.e. ~330 times less) of the DNA template molecules present in an average sample from a BHV bone (Table 2.2.2). Further, only three of the Rosslea DNA extracts contained nuclear DNA at a level allowing amplification with our moa-specific microsatellite primers. The statistics presented in Table 2.2.2 also show that in general, DNA quality in PV fossils was inferior to that of BHV material, even when measured within the same geologic timeframe. In addition to the average C_T values, estimated DNA half-life also differed between the sites although the non-significant R^2 value at BHV prevented a meaningful assessment of molecular half-life at this site. The differences in decay pattern confirmed that age is only one factor determining the DNA quality of moa fossils from North Canterbury. Our results indicate that whatever other factors are present, they seem to be responsible for a faster DNA degradation at Rosslea than at the other sites. We emphasize that the reported $t_{1/2}$ values (Table 2.2.2) should be regarded as estimated means only, and that they strictly apply only to the mtDNA CR fragment targeted here, under the given environmental conditions. Yet, the contention of a relatively fast degradation at Rosslea was supported by the observation that cortical bone of fossils from Rosslea had less structural integrity. It could be a simple effect of age, but equally old moa bones (from other North Canterbury sites) lack these characteristics (personal observation).

Given that the geographical context effectively eliminates climatic conditions as mediators for the observed preservation differences, sediment chemistry is a plausible candidate. For example, Lindahl and Nyberg (1972) showed that with each pH-unit below neutrality, depurination (from DNA in a

solution) increased with an order of magnitude. Despite containing peaty sediments, expected to be acidic, pH at all three sites was almost neutral (PV = 6.5-6.8, BHV = 6.7, Rosslea = 7.2-7.6). Rosslea sediments were apparently weakly alkaline, but pH was measured in preserved sediment samples, which may have influenced the results. The apparent neutrality is likely reflecting the presence of calcareous rock in the area (Wood et al., 2008), which buffers the soils. Obviously, more research is needed to understand the mechanisms responsible for the observed differences.

The significant difference in molecular preservation between bones from PV and those from BHV (Table 2.2.2), despite the sites being contemporaneous, is perhaps easier to explain. BHV was excavated over a short period of time in mid-2001, whereas the PV material was retrieved during several excavations, periodically from late 1930s to 1973 (summarized in Holdaway, Worthy, 1997). So the preservation difference could be a post-excavation storage effect, resulting from an increased rate of DNA damage when fossils have been removed from their depositional environment (Pruvost et al., 2007). Other results suggest that storage time was not a major determinant of DNA preservation in this study. In particular, all Rosslea DNA-samples were obtained within 14 days of the excavation and they still yielded much lower amounts of DNA than material from the other sites. Further, although BHV was excavated only once in 2001, material from this site was more variable in DNA preservation of contemporary samples than were PV fossils, which were excavated and deposited at Canterbury Museum over >40 years.

Lastly, even though an average PV fossil contained just 15 % of the mitochondrial DNA template molecules present in an average BHV fossil, it did not affect the ability to amplify short nuclear DNA fragments (Table 2.2.2). This indicates that nuclear DNA work on moa is not necessarily limited to only the very best specimens. It also confirms a previously-outlined potential for conducting a high resolution microsatellite investigation of late Holocene moa populations (Allentoft et al., 2009).

2.2.6 Concluding remarks

With this study we wish to present Rosslea: a “new” megafaunal fossil site on New Zealand’s South Island. The excavated material from Rosslea is likely to represent only a small fraction of the contents of the deposit. The genetic species identifications of the recovered material matched the morphological assessment and demonstrated the presence of the three most common moa species, of the late Holocene, North Canterbury fossil fauna. Although there were no taxonomic surprises, the Rosslea collection adds another piece in the puzzle towards understanding one of the last ecosystems on the planet to be invaded by humans.

The inclusion of the Rosslea fossils allowed for a much more detailed assessment of ancient DNA degradation through time than has been achieved before. We demonstrate here, an exponential degradation of DNA, probably determined by the depurination rate followed by strand breakage. The considerable variation observed even between contemporary samples from the same site, indicates why many previous studies with more limited data, have failed to demonstrate this association between preservation and time. The three fossil assemblages displayed different patterns in molecular decay, which is interesting in light of their proximity in space and time. The results suggest that a simple model of temperature-dependent DNA degradation is insufficient to describe ancient DNA quality, at least when comparing material recovered within temperate regions. The preservation of DNA in the Rosslea material proved relatively poor, although the moa bones were sampled soon after excavation, which suggests that fossils from such sites should be sampled as soon as possible after recovery. Further degradation in storage might result in the remaining DNA being fragmented beyond accessibility.

We have discussed several hypotheses to explain the observed differences in DNA degradation, but the situation appears to be complex. A thorough study of the physical and chemical sediment properties at each site would be required to resolve this issue. It seems unquestionable, however, that the North Canterbury fossil sites hold the potential for a highly informative comparative investigation, offering unique insight into the factors affecting long-term preservation of DNA in fossils and sediments.

2.2.7 References

- Allentoft ME, Bunce M, Scofield RP, Hale ML, Holdaway RN (2010) Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand's extinct megafauna. *Quaternary Science Reviews* **29**, 753-762.
- Allentoft ME, Schuster S, Holdaway RN, et al. (2009) Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *Biotechniques* **46**, 195-200.
- Bar W, Kratzer A, Machler M, Schmid W (1988) Postmortem stability of DNA. *Forensic Science International* **39**, 59-70.
- Barnett R, Shapiro B, Barnes I, et al. (2009) Phylogeography of lions (*Panthera leo* ssp.) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. *Molecular Ecology* **18**, 1668-1677.
- Bramanti B, Thomas MG, Haak W, et al. (2009) Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* **326**, 137-140.
- Bunce M, Worthy TH, Ford T, et al. (2003) Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425**, 172-175.

- Bunce M, Worthy TH, Phillips MJ, et al. (2009) The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20646-20651.
- Colson IB, Bailey JF (1997) The preservation of ancient DNA and bone diagenesis. *Ancient Biomolecules* **1**, 109-117.
- Cooper A, Lalueza-Fox C, Anderson S, et al. (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704-707.
- Deagle BE, Eveson JP, Jarman SN (2006) Quantification of damage in DNA recovered from highly degraded samples--a case study on DNA in faeces. *Front Zool* **3**, 11.
- Debruyne R, Chu G, King CE, et al. (2008) Out of America: Ancient DNA Evidence for a New World Origin of Late Quaternary Woolly Mammoths. *Current Biology* **18**, 1320-1326.
- Dobberstein RC, Huppertz J, von Wurmb-Schwark N, Ritz-Timme S (2008) Degradation of biomolecules in artificially and naturally aged teeth: Implications for age estimation based on aspartic acid racemization and DNA analysis. *Forensic Science International* **179**, 181-191.
- Gilbert MTP, Binladen J, Miller W, et al. (2007) Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Research* **35**, 1-10.
- Gilbert MTP, Hansen AJ, Willerslev E, et al. (2003) Characterization of genetic miscoding lesions caused by postmortem damage. *American Journal of Human Genetics* **72**, 48-61.
- Gilbert MTP, Jenkins DL, Gotherstrom A, et al. (2008) DNA from pre-Clovis human coprolites in Oregon, North America. *Science* **320**, 786-789.
- Hagelberg E, Bell LS, Allen T, et al. (1991) Analysis of ancient bone DNA - techniques and applications. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **333**, 399-407.
- Hansen AJ, Mitchell DL, Wiuf C, et al. (2006) Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics* **173**, 1175-1179.
- Hansen AJ, Willerslev E, Wiuf C, Mourier T, Arctander P (2001) Statistical evidence for miscoding lesions in ancient DNA templates. *Molecular Biology and Evolution* **18**, 262-265.
- Haynes S, Searle JB, Bertman A, Dobney KM (2002) Bone preservation and ancient DNA: The application of screening methods for predicting DNA survival. *Journal of Archaeological Science* **29**, 585-592.
- Hebsgaard MB, Gilbert MTP, Arneborg J, et al. (2009) 'The Farm Beneath the Sand' - an archaeological case study on ancient 'dirt' DNA. *Antiquity* **83**, 430-444.
- Higham T, Anderson A, Jacomb C (1999) Dating the first New Zealanders: The chronology of Wairau Bar. *Antiquity* **73**, 420-427.
- Higuchi R, Bowman B, Freiburger M, Ryder OA, Wilson AC (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature* **312**, 282-284.
- Holdaway RN, Beavan NR (1999) Reliable C-14 AMS dates on bird and Pacific rat *Rattus exulans* bone gelatin, from a CaCO₃-rich deposit. *Journal of the Royal Society of New Zealand* **29**, 185-211.
- Holdaway RN, Jacomb C (2000) Rapid extinction of the moas (Aves : Dinorinithiformes): Model, test, and implications. *Science* **287**, 2250-2254.
- Holdaway RN, Worthy TH (1997) A reappraisal of the late quaternary fossil vertebrates of Pyramid Valley Swamp, North Canterbury, New Zealand. *New Zealand Journal of Zoology* **24**, 69-121.
- Hoss M, Jaruga P, Zastawny TH, Dizdaroglu M, Paabo S (1996) DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Research* **24**, 1304-1307.
- Huynen L, Millar CD, Scofield RP, Lambert DM (2003) Nuclear DNA sequences detect species limits in ancient moa. *Nature* **425**, 175-178.

- Kaiser C, Bachmeier B, Conrad C, et al. (2008) Molecular study of time dependent changes in DNA stability in soil buried skeletal residues. *Forensic Science International* **177**, 32-36.
- Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715.
- Lindahl T, Andersson A (1972) Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. *Biochemistry* **11**, 3618-&.
- Lindahl T, Nyberg B (1972) Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **11**, 3610-&.
- Lorenz MG, Wackernagel W (1994) Bacterial gene transfer by natural genetic transformation in the environment. *Microbiological Reviews* **58**, 563-602.
- Marota I, Basile C, Ubaldi M, Rollo F (2002) DNA decay rate in papyri and human remains from Egyptian archaeological sites. *American Journal of Physical Anthropology* **117**, 310-318.
- Olivieri C, Ermini L, Rizzi E, et al. (2010) Characterization of Nucleotide Misincorporation Patterns in the Iceman's Mitochondrial DNA. *PLoS ONE* **5**.
- Oskam CL, Haile J, McLay E, et al. (2010) Fossil avian eggshell preserves ancient DNA. *Proceedings of the Royal Society B: Biological Sciences* **277**, 1991-2000.
- Paabo S (1985a) Molecular cloning of ancient Egyptian mummy DNA. *Nature* **314**, 644-645.
- Paabo S (1985b) Preservation of DNA in ancient Egyptian mummies. *Journal of Archaeological Science* **12**, 411-417.
- Paabo S (1989) Ancient DNA - extraction, characterization, molecular-cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 1939-1943.
- Paabo S, Wilson AC (1991) Miocene DNA sequences - a dream come true? *Current Biology* **1**, 45-46.
- Poinar HN, Hofreiter M, Spaulding WG, et al. (1998) Molecular coproscopy: Dung and diet of the extinct ground sloth *Nothotheriops shastensis*. *Science* **281**, 402-406.
- Poinar HN, Hoss M, Bada JL, Paabo S (1996) Amino acid racemization and the preservation of ancient DNA. *Science* **272**, 864-866.
- Pruvost M, Schwarz R, Correia VB, et al. (2007) Freshly excavated fossil bones are best for amplification of ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 739-744.
- Schwarz C, Debruyne R, Kuch M, et al. (2009) New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Research* **37**, 3215-3229.
- Smith CI, Chamberlain AT, Riley MS, Stringer C, Collins MJ (2003) The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of Human Evolution* **45**, 203-217.
- Wood JR, Worthy TH, Rawlence NJ, Jones SM, Read SE (2008) A deposition mechanism for Holocene miring bone deposits, South Island, New Zealand. *Journal of Taphonomy* **6**, 1-20.
- Worthy TH (1987) Sexual dimorphism and temporal variation in the North Island moa species *Euryapteryx curtus* (Owen) and *Pachyornis mappini* (Archy). *National Museum of New Zealand Records* **3**, 59-70.
- Worthy TH, Holdaway RN (1996) Quaternary fossil faunas, overlapping taphonomies, and palaeofaunal reconstruction in north Canterbury, South Island, New Zealand. *Journal of the Royal Society of New Zealand* **26**, 275-361.
- Worthy TH, Holdaway RN (2002) *The lost world of the Moa* Canterbury University Press, Christchurch.

Chapter 3: Identifying microsatellites in the moa genome

Introducing the manuscripts in Chapter 3

The two manuscripts in this chapter introduce microsatellite-based genetic research to the field of ancient DNA. The concept of utilising microsatellite technology in this project was appealing because of the high resolution power offered by these markers, making them ideal to study genetics at the population level.

The first manuscript “Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data” has already been published [Biotechniques (2009), 46, 195-200]. The published paper is found in Appendix 3 and the electronic version in Digital Appendix 19. This study documents the applicability of high-throughput sequencing technology to identify microsatellites. The approach was conceptualised and tested because traditional enrichment methods for identification of repetitive regions failed on the ancient degraded moa DNA. However, the new method has proven cost-benefit efficient on modern DNA as well, and seems to be the method of choice now for microsatellite identification (Google Scholar lists 13 citations of our paper as at 26/8-2010). It is a “proof of concept” paper, describing the entire process from DNA extraction to the GS-FLX 454-run, microsatellite identification, primer design and tests for genetic variability. Lastly, several authentication criteria are established to ensure that the identified microsatellite is indeed of moa origin and not a contaminant picked up by the non targeted “shotgun” sequencing approach.

The next step in the process towards a true aDNA-based microsatellite population study involves the identification of sufficient markers to make such a study viable. The second (unpublished) manuscript in this chapter represents this step, by bridging the proof of concept paper (introduced above) and the actual microsatellite study in Chapter 5. Several high-throughput runs are carried out on aDNA extracts from moa bone as well as moa eggshell, and numerous primers are tested and optimised to identify a sufficient set of polymorphic moa microsatellite loci. It is well known, that amplifying microsatellite loci from low copy number DNA can be problematic because of allelic dropout. To thoroughly investigate this issue, an assessment of dropout is carried out, based on a large number of DNA extracts and multiple PCR repeats. With the results, it is possible to establish a set of criteria and strict validation protocols, ensuring high quality data, despite the fragmented state of the template molecules.

With the methods presented in these two manuscripts, the road is now paved for the first high resolution population study in aDNA research (Chapter 5).

My role has been to obtain and process the samples, to conduct the laboratory work (except the 454 runs) and the data analyses and to write the manuscripts. The first manuscript was conceptualised and written together with Mike Bunce. The GS-FLX 454-runs were conducted by Stephan Schuster, Tom

Gilbert, Morten Rasmussen and Eske Willerslev. Charlotte Oskam extracted eggshell DNA for the 454-runs in the second manuscript. All co-authors have contributed to the interpretations and the writing of the published manuscript.

Identification of microsatellites from an extinct moa species using high throughput (454) sequence data

Morten E. Allentoft ^{1,4}, Stephan C. Schuster ², Richard N. Holdaway ³, Marie L. Hale ¹, Emma McLay ⁴, Charlotte Oskam ⁴, M. Thomas P. Gilbert ⁵, Peter Spencer ⁴, Eske Willerslev ⁵ and Michael Bunce ⁴

1. School of Biological Sciences, University of Canterbury, Christchurch, New Zealand.

2. Pennsylvania State University, Center for Comparative Genomics and Bioinformatics, 310 Wartik Building, University Park, PA 16802, USA.

3. Palaecol Research Ltd, P.O. Box 16 569, Christchurch, New Zealand.

4. Ancient DNA laboratory, School of Biological Sciences and Biotechnology, Murdoch University, Perth 6150, Australia.

5. Department of Biology, University of Copenhagen, Universitetsparken 15, DK2100 Copenhagen, Denmark.

3.1.1 Abstract

Genetic variation in microsatellites is rarely examined in the field of ancient DNA due to the low quantity of nuclear DNA in the fossil record together with the lack of characterised nuclear markers in extinct species. 454 sequencing platforms provide a new high-throughput technology capable of generating up to 1 gigabases per run as short (200-400bp) read-lengths. 454 data were generated from the fossil bone of an extinct New Zealand moa (Aves: Dinornithiformes). We identified numerous short tandem repeat (STR) motifs, and here present the successful isolation and characterisation of one polymorphic microsatellite (Moa_MS2). Primers designed to flank this locus amplified all three moa species tested here. The presented method proved to be a fast and efficient way of identifying microsatellite markers in ancient DNA templates and, depending on biomolecule preservation, has the potential of enabling high-resolution population genetic studies of extinct taxa. As sequence read lengths of the 454 platform and its competitors (e.g., the SOLEXA and SOLiD platforms) increase, this approach will become increasingly powerful in identifying microsatellites in extinct (and extant) organisms, and will afford new opportunities to study past biodiversity and extinction processes.

3.1.2 Introduction

With the introduction of new high-throughput DNA sequencing techniques, capable of generating millions/billions of sequence reads per run, genomic research is advancing faster than ever (Blow, 2008; Mardis, 2008; Medini et al., 2008; Schuster, 2008). In the field of paleogenetics, the first

complete nuclear genome has yet to be recovered, but major sequencing projects of woolly mammoth (Poinar et al., 2006) and Neanderthal (Green et al., 2006) are heading what has been called the “third wave” of progress in ancient DNA (aDNA) (Millar et al., 2008). These new sequencing platforms generate large quantities of protein coding data, which will undoubtedly assist in the study of molecular evolution, functional genomics and adaptation. In addition to coding data, the huge number of randomly amplified sequences provides the opportunity to search for microsatellites or short tandem repeats (STRs). These non-coding sequences, with a high rate of mutation, are applied as markers in a wide array of genetic research, especially in relation to forensics, modern population biology, and parentage analyses.

A limited number of successful STR amplifications have been reported from ancient substrates using “modern” STRs as templates, e.g. (Amory et al., 2007; Greenwood et al., 2001; Manen et al., 2003) but to our knowledge, no microsatellite primers developed directly from aDNA templates (let alone extinct species) have been published prior to this study. This is likely due to complications in traditional microsatellite library construction as the result of the degraded and cross-linked nature of aDNA (Paabo, 1989). Greenwood et al. (Greenwood et al., 2001) managed to amplify a single microsatellite locus in woolly mammoth using primers developed for modern elephants. However, to rely solely on primers developed for related modern taxa, when targeting microsatellites in extinct ones, is problematic because of possible low cross-species amplification rates and chance of monomorphism in the target species (Primmer et al., 1996; Primmer et al., 2005). This is especially pertinent in taxa such as moa where the likelihood of cross-species amplification is limited as a result of the >80 million years that separate moa from their closest living relatives among the ratite birds (Cooper et al., 2001)². Consequently, the chance of identifying polymorphic microsatellite markers in ancient DNA templates of an extinct taxon seems greatly enhanced when the potential markers have been identified directly on the target species.

The Roche GS-FLX (454 Life Sciences, Branford, CT, USA) sequencing technology is currently capable of producing 0.1 gigabases per run with a read length averaging 200-300 nucleotides – a sequence size that allows for the presence of an STR and sufficient flanking regions to design primers. A GS-FLX run was conducted on a *Pachyornis elephantopus* (heavy-footed moa) bone extract to identify a series of microsatellite loci. To illustrate the viability of this technique, we identified an (AC)₁₂ microsatellite (directly from “raw” GS-FLX data) and demonstrated cross-species amplification in three species of moa. All eleven New Zealand moa species³ were driven to extinction

² Since the publication of this paper, Phillips et al. (2010) have estimated that moa divergence did not happen before ~60 Mya (see Table 1.1.1)

³ Since the publication of this paper, Bunce et al. (2009) revised the taxonomy and the present consensus is nine moa species.

in the early 15th century, following the arrival of Polynesians. Identification of new STR markers, such as described here, will enable detailed DNA profiling of extinction processes and past population dynamics of these ancient ratites.

3.1.3 Materials and methods

3.1.3.1 Sampling of moa fossils and DNA isolation

Sampling of moa fossils was conducted by drilling out cylindrical elements (diameter ~ 1 cm) of moa tibiotarsal bones using a power drill and diamond-dust coated drill-bits. Each sample was then ground into bone powder using a Dremel tool (Part no. 114; Racine, WI, USA). To minimize the incorporation of any possible DNA contamination present on the bones, the bone surfaces and the inner porous parts were excluded, and only solid cortical bone was processed. Contamination from external sources, as well as cross-contamination between samples, was minimised by thorough cleaning equipment and sampling environment (with 10% bleach and 100% alcohol) between the processing of each individual. To minimize the risk of a ubiquitous DNA contaminant being present on all the bones, fossils representing three different moa species, from two different sites, and from several different museum/university collections were included (Digital Appendix 7). Pyramid Valley and Bell Hill Vineyard Swamp both represent late Holocene deposits in North Canterbury, New Zealand with a known fossil record spanning app. 3700-700 BP [(Burrows, 1989), RNH – unpub data]⁴.

DNA was extracted from 200 mg of bone powder through incubation with rotation at 55°C for 48 hours in 1.5 ml digestion buffer [20 mM Tris, pH 8.0, 1% Triton X-100, 10mM dithiothreitol (DTT), 1 mg/ml proteinase K and 0.5 M EDTA]. The supernatant was spun through Centricon 30,000 molecular weight cut-off (MWCO) (Millipore, Billerica, MA, USA) columns combined with 5 volumes of PBI buffer (Qiagen, Valencia, CA) and DNA was then extracted using silica spin columns (Qiagen) and cleaned with AW1 and AW2 wash buffers (Qiagen) before final elution in 60 µl of 10mM Tris buffer. For species identification, and to confirm the relative quantities of aDNA in the fossil extracts, a 242-bp sequence of moa mitochondrial control region was amplified (by qPCR) using the species-specific primer set (262f/441r) (Digital Appendix 8) and then sequenced as described in Bunce et al. (2003) (data not shown).

⁴ Radiocarbon data are found in Chapter 4 and Digital Appendix 2

3.1.3.2 GS-FLX sequencing, microsatellite primer development and PCR

A DNA extract from a single moa individual (*P. elephantopus*) from the Bell Hill Vineyard Swamp, which demonstrated good qPCR amplification and long nuclear amplicons (data not shown), was picked for a quarter-plate run on the GS-FLX instrument (Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, PA, USA). The moa DNA library was constructed, as previously described (Margulies et al., 2005; Poinar et al., 2006), with the modification of leaving the extracted DNA unsheared before blunt-ending and phosphorylating the DNA fragments by enzymatic polishing using T4 DNA polymerase, T4 polynucleotide kinase, and *E. coli* DNA polymerase. The blunt-ended, double-stranded DNA fragments were then subjected to adapter ligation followed by isolation of the single-stranded template DNA (sstDNA). Subsequently, DNA library fragments were captured onto beads and clonally amplified within individual emulsion droplets (emPCR). The emulsions were disrupted using isopropanol and beads containing amplified DNA fragments were enriched and recovered for sequencing. The recovered sstDNA beads were packed onto a quarter division of a 70mm x 75mm PicoTiterPlate (454 Life Sciences) and loaded onto the GS-FLX sequencing system as previously described (Poinar et al., 2006). The sequencing run yielded 79,796 sequences averaging 112-bp in length. Subsequently, the data were screened for STR sequences using MSATCOMMANDER (Faircloth, 2008) and a total of 195 di-, tri-, and tetra-nucleotide repeat sequences were detected (23 of them with six or more repeats). Seven of these proved particularly promising, with flanking regions of sufficient lengths and a base composition suitable for primer design (Figure 3.1.1 and Digital Appendix 9). The sequence chosen for initial trial was clone # 103234_2765_0456, a 158-bp sequence which included an (AC)₁₂ dinucleotide repeat (Figure 3.1.1, Genbank accession no. FJ513189). Primers were then manually designed from this original clone to target a 114-bp sequence, designated Moa_MS2 (Figure 3.1.1).

Each microsatellite PCR was conducted in a 25- μ l volume containing 2 μ l DNA extract, 1 μ l 10 mg/ml bovine serum albumin (BSA), 1 each of μ l 10 mM forward (5'-TGAGCACCAATACAACCTTCATGG-3') and reverse primer (5'-GACTGTTATTCTATTCCAGTATATGTTTG-3'), 2.5 μ l 10 x ABI Gold Buffer, 2.5 μ l 25mM MgCl₂, 0.2 μ l 5 units/ μ l ABI Gold Taq polymerase, 0.25 μ l of each dNTP (25 mM) and 14.55 μ l Ultrapure H₂O (Catalog nos. 4311816 and AM 9935; Applied Biosystems, Foster City, CA, USA). Thermal cycling was initiated with a 3-min 95° denaturation step followed by 50 cycles of 95° for 20 s, 57° for 45 s, and 72° for 45 s. A final extension was carried out at 72° for 10 min to maximize adenylation. Negative extraction and amplification controls were always included. DNA fragments were separated on an ABI 3730 genetic analyser and sized by co-running a Genescan LIZ500 size standard (Applied Biosystems, Foster City, CA). DNA fragments were scored manually with the aid of GENEMARKER v 1.5 (Soft Genetics, State College, PA, USA). To check the integrity of the

microsatellite, several individual Moa_MS2 amplification products were cloned using a TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA) and sequenced using vector specific M13 primers according to the manufacturer's instructions. Comparisons of the clone sequence data with the original GS-FLX clone confirmed that the Moa_MS2 locus was amplified (data not shown). In accordance with aDNA guidelines, DNA extractions and PCR setup occurred in separate and dedicated aDNA facilities at Murdoch University (Perth, Australia).

3.1.4 Results and discussion

Of the 74 bone DNA extracts prescreened with mtDNA control region amplifications, 52 amplified with the Moa_MS2 primers. The generated chromatographs were consistent with expected dinucleotide STR patterns (including stutters and A+ peaks, see Figure 3.1.1 and Digital Appendix 10). The Moa_MS2 locus yielded a total of 13 different alleles in the three extinct moa genera tested here (*Dinornis*, *Pachyornis* and *Euryapteryx*). Summary statistics for this single locus are presented in Table 3.1.1. Because of the relatively low quality of the template molecules in the fossils, some alleles did not always amplify with the same efficacy in heterozygotes and allele dropout was observed on occasion. This is not totally unexpected for microsatellites in ancient or low copy number substrates (Burger et al., 1999), and when this marker is applied in actual population research (M.A., R.N.H., and M.B., unpublished data), these issues would need to be accounted for through multiple replicates and established protocols (Morin et al., 2001; Taberlet et al., 1996). The chromatographs shown in Figure 3.1.1 are a sample of profiles (two homozygotes and two heterozygotes) obtained from 4 moa fossils and demonstrate the 2-bp stuttering characteristic of dinucleotide STR in addition to A+ peaks. Other chromatographs can be found in the supplementary information (Digital Appendix 10).

The developed primer set amplified across all the three moa genera tested, and more importantly, displayed polymorphism (Table 3.1.1), making it ideal for population studies. Attempts to amplify the Moa_MS2 locus in other ratites (emu, ostrich, rhea) were unsuccessful but considering the deep split to the most recent common ancestor, this was not unexpected. Since this study serves to document the use of high-throughput sequence data as a source of microsatellites, a further characterization of the allelic diversity in relation to the various moa species is beyond the scope of this proof-of-concept paper. Clearly, population genetic analyses based on microsatellite data require multiple polymorphic loci. Significantly, our results demonstrate how several potential markers were identified in this quarter-plate GS-FLX run. Although only one STR has been evaluated here, at least seven STRs appeared potentially suitable as genetic markers (Digital Appendix 9). Given that birds exhibit a relatively low frequency of microsatellites (Primmer et al., 1997), the frequency of detected

microsatellites per run may be even greater in other vertebrate lineages. We screened ~64,000 clones of Genbank mammoth data (Poinar et al., 2006), generated using GS-20 sequencing platform (454 Life Sciences), for microsatellites and identified 15 STRs with ≥ 6 repeating units (data not shown). The shorter read lengths of the GS-20 (~100 bp), used in the mammoth study, likely explains the slightly lower frequency of identified microsatellites (0.023%) when compared to the GS-FLX (~200bp) moa run (0.029%).

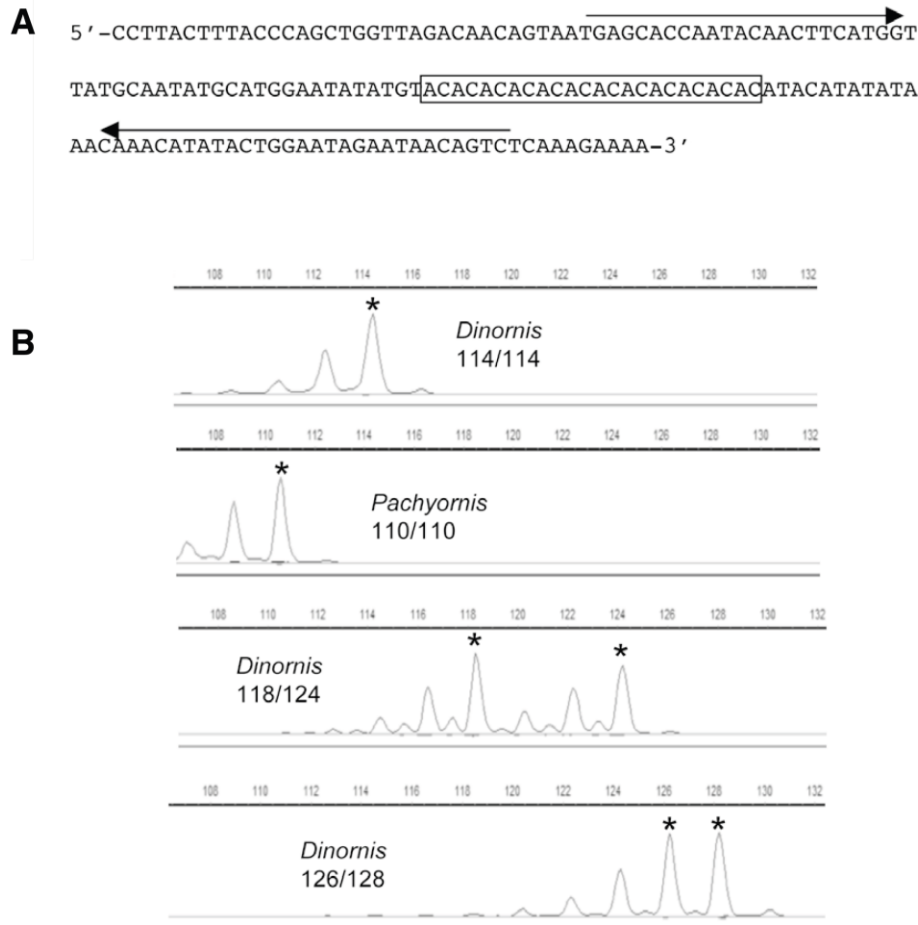


Figure 3.1.1: A) The original clone identified in the GS-FLX data (clone # 103234_2765_0456). The (AC)₁₂ repetitive region is boxed and the location of forward and reverse primer sequences indicated by arrows. B) Chromatographs obtained from the amplification of the Moa_MS2 locus in two moa genera (*Dinornis* and *Pachyornis*). Modified screenshot from GENEMARKER v 1.5. The upper two panels show homozygotes and the two lower panels show heterozygotes. Alleles are marked with an astrisk and additional peaks represent stutter and A+ peaks, commonly observed during STR profiling. Chromatographs from other species/individuals are found in Digital Appendix 10.

The characterisation of DNA from the fossil record is not without challenges. In fossil bones recovered from soil, a fraction of the present biomolecules is likely to be of bacterial or fungal origin,

see (Noonan et al., 2005; Poinar et al., 2006). Moreover, museum specimens are commonly contaminated with human DNA from previous handling (Gilbert et al., 2005). When using “shotgun” sequencing methods such as described here, all DNA molecules in an extract are amplified with equal chance (however, see reference (However, see reference Dohm et al., 2008), including any exogenous DNA. Hence, in theory, there is a potential risk of identifying microsatellites – and even conduct population analyses – on the microbial flora/fauna or other contaminants associated with the fossils instead of the target species itself. With regard to the data presented here, we consider human contamination unlikely because the flanking regions of Moa_MS2 demonstrated no significant matches when queried against the human genome database. The paucity of environmental microbe-sequences in GenBank makes a similar approach to detect microbes more problematic, but in the case of Moa_MS2, it is unlikely to be of microbial origin for the following reasons: (i) Though present, microsatellites are much scarcer in prokaryotes and fungus compared with higher eukaryotes (Field, Wills, 1998; Karaoglu et al., 2005; Schlotterer et al., 2006; Toth et al., 2000). (ii) Many heterozygotes were detected among the 52 genotyped moa, whereas the haploid nature of bacteria should generate only one allele per clone. (iii) Fungi exist in a variety of ploidy states, and we never observed more than 2 alleles in any of the fossil bone profiles. (iv) Fossils from different collections and originating from very different depositional environments in New Zealand, with different microbial communities, still yielded reproducible amplifications. (v) Allele frequencies differed among species (Digital Appendix 11); for example, alleles 106, 114 and 124 are considered private alleles for *Pachyornis*, *Dinornis* and *Euryapteryx* respectively. These findings are incompatible with the possibility that the locus is not of moa origin. (vi) Lastly, the qPCR data presented in Digital Appendix 8 show a significant difference (t -test, $P = 0.000038$) between mean mtDNA Control region C_T values for those DNA extracts that yielded a Moa_MS2 profile (mean $C_T = 32.4$) and those that yielded mtDNA but failed to amplify the Moa_MS2 locus (mean $C_T = 36.7$). Fossils with poor mtDNA preservation (high C_T values) had a lower success rate for amplification of the nuclear marker. A quantitative correlation, such as observed here, should not exist if the microsatellite originated from a source other than moa.

In summary, data generated using a high-throughput DNA sequencing technique were pivotal in identifying a polymorphic microsatellite marker in the extinct New Zealand moa. Mining genomic data for microsatellites is especially relevant to ancient DNA where the preservation is not conducive to more traditional approaches of library enrichment and construction. We argue that STRs identified using “shotgun” sequencing approaches, coupled with better DNA isolation techniques, will facilitate research into high-resolution ancient population genetics similar to those currently conducted on extant species.

Table 3.1.1: Characteristics of the microsatellite locus Moa_MS2 that was optimized for use in three extinct moa species, including the South Island giant moa (*D. robustus*), stout-legged moa (*E. gravis*)⁵ and the heavy-footed moa (*P. elephantopus*). Descriptive measures include n , the number of individuals; N_A , number of alleles; N_E , effective number of alleles; H_O , observed and H_E , expected heterozygosity and F_{IS} , Wright's fixation index as a measure of heterozygote deficiency/excess.

Taxon	Size range (bp)	n	N_A	N_E	F_{IS}	Heterozygosity	
						H_O	H_E
<i>D. robustus</i>	110-136	31	10	3.1	0.143	0.581	0.677 ± 0.089
<i>E. gravis</i>	110-128	11	8	5.8	0.23	0.636	0.826 ± 0.145
<i>P. elephantopus</i>	106-110	10	2	1.9	0.167	0.4	0.505 ± 0.155

As average 454 read lengths increase to ~400 bp (Droege, Hill, 2008) and the costs of sequencing decrease, the number of potential microsatellite markers identified per run can increase dramatically. It may not be long before shotgun sequencing will replace traditional enrichment library approaches, with one major benefit being the large amount of ancillary mitochondrial and nuclear data that are generated. Moreover, when enrichments steps are combined with coded primers (Binladen et al., 2007), it will be possible to generate microsatellite libraries from multiple species simultaneously. In the context of aDNA, the use of enrichment steps may represent a number of additional technical challenges (such as extent of cross linking) and may be dependent on DNA preservation within the substrate.

3.1.5 Acknowledgements:

We would like to gratefully acknowledge the following institutions for samples: Museum of New Zealand Te Papa Tongarewa (A.J.D Tennyson), Canterbury Museum (P. Scofield) and The American Museum of Natural History. We thank Malene Møhl, Fiona Wilson, James Haile and Ross Barnett for extensive help during the bone sampling process and Lynn Tomsho for technical assistance in the GS-FLX sequencing. Financial support was provided by The Australian Research Council (DP0771971) and the Marsden Fund of the Royal Society of New Zealand (programme 06-PAL-001-EEB) with Palaecol Research Ltd. Stephan C. Schuster is supported in part by the Gordon and Betty Moore Foundation.

⁵ Since the publication of this paper, Bunce et al. (2009) has revised the taxonomy and *Euryapteryx gravis* is now called *Euryapteryx curtus* – which is the name applied in the other chapters.

3.1.6 References

- Amory S, Keyser C, Crubezy E, Ludes B (2007) STR typing of ancient DNA extracted from hair shafts of Siberian mummies. *Forensic Science International* **166**, 218-229.
- Binladen J, Gilbert MTP, Bollback JP, et al. (2007) The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. *PLoS ONE* **2**, e197.
- Blow N (2008) DNA sequencing: generation next-next. *Nature Methods* **5**, 267-+.
- Bunce M, Worthy TH, Ford T, et al. (2003) Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425**, 172-175.
- Bunce M, Worthy TH, Phillips MJ, et al. (2009) The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20646-20651.
- Burger J, Hummel S, Herrmann B, Henke W (1999) DNA preservation: A microsatellite-DNA study on ancient skeletal remains. *Electrophoresis* **20**, 1722-1728.
- Burrows CJ (1989) Moa browsing - evidence from the Pyramid Valley mire. *New Zealand Journal of Ecology* **12**, 51-56.
- Cooper A, Lalueza-Fox C, Anderson S, et al. (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704-707.
- Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucl. Acids Res.* **36**, e105-.
- Droege M, Hill B (2008) The Genome Sequencer FLX(TM) System--Longer reads, more applications, straight forward bioinformatics and more complete data sets. *Journal of Biotechnology* **136**, 3-10.
- Faircloth BC (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* **8**, 92-94.
- Field D, Wills C (1998) Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S-cerevisiae*, result from strong mutation pressures and a variety of selective forces. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1647-1652.
- Gilbert MTP, Rudbeck L, Willerslev E, et al. (2005) Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy. *Journal of Archaeological Science* **32**, 785-793.
- Green RE, Krause J, Ptak SE, et al. (2006) Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**, 330-336.
- Greenwood AD, Lee F, Capelli C, et al. (2001) Evolution of endogenous retrovirus-like elements of the woolly mammoth (*Mammuthus primigenius*) and its relatives. *Molecular Biology and Evolution* **18**, 840-847.
- Karaoglu H, Lee CMY, Meyer W (2005) Survey of simple sequence repeats in completed fungal genomes. *Molecular Biology and Evolution* **22**, 639-649.
- Manen JF, Bouby L, Dalnoki O, et al. (2003) Microsatellites from archaeological *Vitis vinifera* seeds allow a tentative assignment of the geographical origin of ancient cultivars. *Journal of Archaeological Science* **30**, 721-729.
- Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**, 133-141.
- Margulies M, Egholm M, Altman W, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380.

- Medini D, Serruto D, Parkhill J, et al. (2008) Microbiology in the post-genomic era. *Nature Reviews Microbiology* **6**, 419-430.
- Millar CD, Huynen L, Subramanian S, Mohandesan E, Lambert DM (2008) New developments in ancient genomics. *Trends in Ecology & Evolution* **23**, 386-393.
- Morin PA, Chambers KE, Boesch C, Vigilant L (2001) Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology* **10**, 1835-1844.
- Noonan JP, Hofreiter M, Smith D, et al. (2005) Genomic sequencing of Pleistocene cave bears. *Science* **309**, 597-600.
- Paabo S (1989) Ancient DNA - extraction, characterization, molecular-cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 1939-1943.
- Poinar HN, Schwarz C, Qi J, et al. (2006) Metagenomics to paleogenomics: Large-scale sequencing of mammoth DNA. *Science* **311**, 392-394.
- Primmer CR, Moller AP, Ellegren H (1996) A wide-range survey of cross-species microsatellite amplification in birds. *Molecular Ecology* **5**, 365-378.
- Primmer CR, Painter JN, Koskinen MT, Palo JU, Merila J (2005) Factors affecting avian cross-species microsatellite amplification. *Journal of Avian Biology* **36**, 348-360.
- Primmer CR, Raudsepp T, Chowdhary BP, Moller AR, Ellegren H (1997) Low frequency of microsatellites in the avian genome. *Genome Research* **7**, 471-482.
- Schlotterer C, Imhof M, Wang H, Nolte V, Harr B (2006) Low abundance of Escherichia coli microsatellites is associated with an extremely low mutation rate. *Journal of Evolutionary Biology* **19**, 1671-1676.
- Schuster SC (2008) Next-generation sequencing transforms today's biology. *Nature Methods* **5**, 16-18.
- Taberlet P, Griffin S, Goossens B, et al. (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**, 3189-3194.
- Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research* **10**, 967-981.

Profiling the dead: a characterization of six ancient DNA microsatellite markers for the extinct New Zealand moa – protocols, problems, and prospects

3.2.1 Abstract

We present the first set of microsatellite markers developed exclusively for an extinct taxon. Microsatellites have been applied in thousands of genetic studies on extant species over the last two decades. Now, with the primers and protocols presented here, the same technology is available to study the extinct moa (Aves: Dinornithiformes) populations of New Zealand. Many microsatellite sequences were identified using high throughput sequencing technology (GS-FLX) on ancient DNA extracted from fossil moa bone and eggshell. From the “shotgun” reads >60 primer pairs were designed and tested on DNA from the South Island giant moa (*Dinornis robustus*) and six polymorphic loci were characterised in the process. Because of the low number of nuclear DNA copies in the fossil record, allelic dropout proved to be severe and affected 36-70% of the PCR reactions per microsatellite marker. However, a comprehensive survey of allelic dropout, combined with supporting quantitative PCR data, allowed us to establish a set of criteria that effectively minimised the impact of this effect. Finally, we demonstrated the viability of the primers and the protocols, by compiling a *Dinornis* microsatellite dataset from 74 individuals, which showed no signs of being compromised by allelic dropout. The methodology presented here provides a framework by which to generate and evaluate microsatellite data from ancient templates and thereby opens a new avenue in ancient DNA research.

3.2.2 Introduction

The discovery and use of polymorphic microsatellite loci (also known as STRs or short tandem repeats) have had a major impact in many areas of genetic research. For the past two decades they have been the markers of choice in a wide range of forensic genetics, population genetics and wildlife-related research. The importance and applicability of these markers are confirmed by observing an excess of 45000 hits on the word “microsatellite” on the Web of Science database (accessed mid-2010). With the invention of high-throughput sequencing platforms, such as the GS-FLX (454 Life Sciences of Roche, Branford, CT, USA), microsatellite marker development has recently become fast and efficient (Abdelkrim et al., 2009; Allentoft et al., 2009; Santana et al., 2009)

and the advance has also allowed the identification of these loci in highly degraded ancient DNA (aDNA), where traditional enrichment procedures can fail (Allentoft et al., 2009).

Here, we present the first set of microsatellite markers to be developed directly from aDNA templates with the specific aim of studying an extinct taxon: the New Zealand moa (Aves: Dinornithiformes). The moa lineage included nine species (Bunce et al., 2009) of very large herbivorous birds (15-250 kg). They occupied most areas within New Zealand until their sudden extinction shortly after Polynesians colonized the landmass in the late 13th century (Holdaway, Jacomb, 2000). However, bones, coprolites, feathers and eggshell of moa are regularly discovered and some natural fossil sites have yielded well-preserved bones in very high concentrations (e.g., Wood et al., 2008; Worthy, Holdaway, 1996; Worthy, Holdaway, 2002). These offer rare, if not unique opportunities to study extinct megafauna at the population level. As demonstrated recently, analyses of well-preserved aDNA from hundreds of moa individuals found in close spatial and temporal proximity can uncover patterns in the biology and local population dynamics almost at the level of research on living populations (Allentoft et al., 2010). However, to fully extract and explore the information within this unique paleontological context, a set of high resolution genetic tools is necessary. Since microsatellites have a high mutation rate and are inherited from both sexes, they can offer more detailed population genetic insight compared to maternally inherited mitochondrial DNA (mtDNA) as traditionally targeted in ancient substrates. A functional set of microsatellite markers would potentially allow research into moa biology, population structure and the extinction process in much greater detail than has been attempted before with aDNA.

The procedure for identifying microsatellites from aDNA has already been established (Allentoft et al., 2009), so the two major challenges in the present study were: (1) to identify and characterize a sufficient number of useable and polymorphic loci to yield informative data and (2) to investigate whether data quality was compromised by having been generated from aDNA templates. Given good DNA preservation, the first issue is essentially a matter of funds and effort to identify, design, and test enough primers. The second issue is, however, potentially more problematic. Ancient DNA research generally targets mitochondrial DNA (mtDNA) because of the many mitochondrial genome copies per cell, whereas aiming at single-copy nuclear markers such as microsatellites is more challenging, and requires stringent methodology and careful evaluation of the results.

The main concern, apart from PCRs failing completely, is “allelic dropout”. Dropout is observed when amplifying low copy number DNA (LCN DNA), since the stochastic nature of the earliest cycles in the PCR reaction can cause one of the two (in diploids) alleles to swamp the PCR, completely suppressing amplification of the other allele. The outcome is a genetic profile resembling that of a homozygote regardless of the true genotype of the sample. Allelic dropout is well known in PCRs with LCN DNA (e.g., Miller et al., 2002; Navidi et al., 1992; Roeder et al., 2009; Taberlet et

al., 1996) and has been observed and assessed in DNA extracts from a variety of substrates, including faeces (Frantz et al., 2003; Morin et al., 2001; Parsons, 2001), hair (Goossens et al., 1998; Morin et al., 2001), feather (Sefc et al., 2003), tooth (Arandjelovic et al., 2009), and fingerprints (Balogh et al., 2003). The magnitude of the allelic dropout can vary significantly between samples, between different primers, and also between alleles within the same microsatellite marker, so a case-by-case assessment is necessary in order to establish the extent of the problem and the required protocols. Most important in this context, is to determine how many independent PCR repeats that are necessary before the genotype of an apparent homozygous individual can be trusted (e.g., Navidi et al., 1992; Taberlet et al., 1996).

For ancient moa DNA, it was apparent very quickly that dropout was affecting the allele scoring (personal observation). Therefore, following the successful identification of each suitable microsatellite, 88 DNA extracts of the South Island giant moa (*Dinornis robustus*) from fossil locations in North Canterbury, were screened. Doing so, allowed us to characterise in detail, the allelic dropout in each marker, and to establish a set of criteria to minimise the effects. To test the viability of the established protocols, a full multi-locus *Dinornis* dataset was assembled and examined thoroughly for signs of allelic dropout, such as a deficit of heterozygous individuals in comparison to the expected Hardy-Weinberg proportions. Concurrently, results of a comprehensive quantitative PCR (qPCR) experiment on the extracts were used to investigate a link between DNA preservation and dropout and also between DNA preservation and individual observed homozygosity.

3.2.3 Materials and methods

3.2.3.1 Sampling and DNA extraction

Detailed descriptions of the samples, bone sampling protocols, DNA extractions, PCR conditions and sequencing for species identification are given in Allentoft et al. (2010; 2009). The moa eggshell DNA used for the GS-FLX run was extracted according to protocols in Oskam et al. (2010). In accordance with aDNA guidelines, all DNA extractions and PCRs reported in this study were performed in a separate and dedicated aDNA laboratory at Murdoch University (Perth, Australia).

3.2.3.2 High throughput sequencing

The Moa_MS2 microsatellite was described in a “proof of concept” study based on data from a high throughput, 454 GS-FLX run (Allentoft et al., 2009). Despite attempts, no more polymorphic microsatellites could be identified from those data, so additional GS-FLX data was generated, from a

total of 10 moa DNA extracts (Table 3.2.1). Six of these extracts were prepared as conventional FLX shotgun libraries and each was sequenced on 1/8th of a PicoTiter plate (454 Life Sciences of Roche, Branford, CT, USA) using LR70 chemistry. The FLX process followed the manufacturer's guidelines with the following exceptions: 1) As the aDNA was likely fragmented *a priori* due to post mortem damage, initial nebulization was omitted. 2) Library release from beads prior to quantification used the heat-release protocol advocated by Maricic and Pääbo (2009). 3) Prior to emPCR, the DNA quantity in each library was determined using qPCR (Meyer et al., 2008).

The remaining four extracts were prepared as MID-tagged libraries with the same as above modifications to the manufacturer's guidelines. Following emPCR, approximately equal quantities of enriched beads from each library were pooled into a single library, and sequenced on a full PicoTiter plate using LR70 chemistry. Prior to data analyses, the sequence reads from this dataset were bioinformatically sorted by MID tag into the original four libraries.

3.2.3.3 Microsatellite primer design and optimisation

Following 454-sequencing, the thousands of recorded reads were screened for repetitive nucleotide regions using MSATCOMMANDER (Faircloth, 2008). The candidate sequences were inspected individually, to choose sequences showing a large number of repeats, as well as sufficient flanking region and base composition suitable for primer design. Microsatellite primers were then designed using the web-based program PRIMER3PLUS (Untergasser et al., 2007), with settings ensuring short PCR products – preferably shorter than 100bp – to maximize their applicability to highly-fragmented ancient DNA. Each primerset was then tested and optimized on DNA extracts from South Island giant moa (*Dinornis robustus*), using the 50-cycle microsatellite PCR protocol presented in Allentoft et al. (2009), but with annealing temperatures varied throughout the respective tests and optimisations. Given the relative difficulty in obtaining samples for ancient DNA research, all PCR reactions were here done in halved volumes (a total of 12.5 µl per reaction rather than 25 µl) and hence using just 1 µl of the undiluted DNA extract per reaction. PCR products were screened with electrophoresis on 2 % agarose gels. If bands were visualised in the expected size-range, primers were re-ordered with fluorescent 6-FAM dye (Integrated DNA Technologies, Coralville, IA, USA) and the PCRs were repeated with these for accurate size determination. DNA fragments were separated on an ABI 3730 genetic analyser and sized with Genescan LIZ500 size standard (Applied Biosystems, Foster City, CA, USA). Negative and positive controls were always included. The alleles were scored using GENEMARKER version 1.5 (Soft Genetics, State College, PA, USA).

3.2.3.4 Summary statistics, HW-tests and allelic dropout

A large allele-call database was established based on more than 1500 successful PCR reactions across the six markers, using 88 available *Dinornis robustus* DNA extracts (representing the same number of individuals, as extracts were derived from 88 different left tibiotarsa bone elements). Each individual was successfully genotyped an average of 2.9 times for each of the six markers. Dropout for each marker was then calculated as the proportion of PCR reactions in known heterozygotes that amplified only one of the two alleles. A heterozygote was simply determined as showing either a clear bi-allelic profile in one PCR, or two different mono-allelic profiles in different PCRs. Hence, an individual showing AA, BB, BB, and AB in four different successful PCRs would have a dropout of 75%, whereas an individual showing AB, AB in two different PCRs would have 0% dropout.

Basic descriptive statistics of the final assembled microsatellite data, included measures of genetic diversity (N_A , N_E , H_O , H_E) and fixation indices (F_{IS}), were calculated using The Microsatellite Toolkit (Park, 2001) and GENALEX 6 (Peakall, Smouse, 2006). Tests for deviations from Hardy-Weinberg proportions within and across the microsatellite markers and linkage disequilibria between markers were performed with the methods implemented in GENEPOP v.4.0.10 (Rousset, 2008) using default settings. MICROCHECKER ver. 2.2.3 (van Oosterhout et al., 2006) was used to further assess the quality of the data, by looking for evidence of scoring errors resulting from stuttering, null alleles, or drop-out of long alleles.

3.2.3.5 Quantitative PCR

A large proportion of our available moa DNA extracts had already been screened (with the moa specific 262F/441R mtDNA Control region primers) in a qPCR amplification setup using SYBR detection chemistry on a Bio-Rad My-IQ thermocycler as described previously (Allentoft et al., 2009; Allentoft et al., 2010). The C_T values were recorded as an accurate proxy for the relative DNA preservation in each extract, in the absence of PCR inhibitors. A high C_T -value reflects a late takeoff in the monitored PCR reaction because of a limited number of suitable template molecules. As outlined further below, this assessment provided a valuable resource when validating the microsatellite data that were generated.

3.2.4 Results

3.2.4.1 High-throughput (GS-FLX) sequencing

Results for the two GS-FLX runs representing ten aDNA extracts, two types of substrate and four moa genera, are shown in Table 3.2.1. In total, 603,773 sequences were generated and 1809 relevant repetitive regions (di- tri- and tetra-nucleotides with six or more repeat units) were detected among those, corresponding to 0.3% of the sequences. The number of sequences in the two different GS-FLX runs cannot be compared because of the different set-ups (i.e. separated lanes contra MID-tagging), but it seems that the eggshell extracts provided longer average read lengths and also a higher fraction of potential STRs. Notably, within run 1, the three eggshell extracts, combined, produced an average of 0.47% STRs and an average read length of 124.6 bp, whereas the same parameters for bone were 0.1% and 107.8 bp.

Table 3.2.1: Summary of output from two GS-FLX runs. Ten moa DNA extracts were used, representing two types of substrate and four moa species. Run 1 was conducted with traditional 454-libraries and each of the six moa extracts occupied 1/8 of a PicoTiter plate. Run 2 was done with pooled MID-tagged libraries added to a full plate. *Reads* and *Average length* describe the total number of filtered sequences and average length. *STRs* is number of identified microsatellites (≥ 6 repeat units) and *Fraction* is number of STRs per outputted sequence.

Taxon	Extract #	Substrate	Run #	Plate	Reads	Av. length	STR's	Fraction
<i>Dinornis robustus</i>	MB146	eggshell	1	1/8	16572	133.4	116	0.0070
<i>Dinornis robustus</i>	MB147	eggshell	1	1/8	21090	122.4	33	0.0016
<i>Dinornis robustus</i>	MB149	eggshell	1	1/8	19956	119.6	119	0.0060
<i>Dinornis robustus</i>	MB550	bone	1	1/8	27829	113.1	72	0.0026
<i>Dinornis robustus</i>	PIBH16A	bone	1	1/8	40007	104.4	25	0.0006
<i>Pachyornis elephantopus</i>	MB551	bone	1	1/8	40447	107.6	9	0.0002
<i>Dinornis robustus</i>	DBHS	eggshell	2	MID	85476	123.7	37	0.0004
<i>Dinornis robustus</i>	MB572	eggshell	2	MID	125254	120.4	566	0.0045
<i>Pachyornis mappini</i>	MB595	eggshell	2	MID	20934	85.5	14	0.0007
<i>Anomalopteryx didiformes</i>	MB718	eggshell	2	MID	206208	119.1	818	0.0040
Total from eggshell in run 1		eggshell	1		57618	124.6	268	0.0047
Total from bone in run 1		bone	1		108283	107.8	106	0.0010
Total from eggshell in run 2		eggshell	2		437872	118.8	1435	0.0033
Overall					603773	115.4	1809	0.0030

3.2.4.2 Primer design and initial testing

Although 1809 repetitive regions were detected, only a small fraction of them proved suitable as potential microsatellite markers. The reasons for this large rejection of candidate sequences were either an unsuitable base composition (for example, too high/low GC content), or more often, because

insufficient flanking region was revealed to fit the primers. We designed and tested 61 primer pairs from the GS-FLX data but also tried primers developed for other ratites (Brennan, Hyseni, 2008; Huang et al., 2008; Kimwele et al., 1998; Tang et al., 2003; Taylor et al., 1999; Ward et al., 1998). In addition, several generic bird microsatellite markers (Dawson et al., 2010) were re-designed (shortened) for aDNA purposes and tested. A total of 89 primer pairs was evaluated in the process, across a range of annealing temperatures in the PCRs. Ultimately, only five proved suitable (plus the already described Moa_MS2) as genetic markers (Table 3.2.2), and met the authentication criteria established in Allentoft et al. (2009). The rest either failed to amplify anything, amplified the wrong fragments, or amplified the right fragments but were not polymorphic. For the latter condition, we note that all the DNA extracts were from individuals excavated from the same area (North Canterbury, South Island); some of these apparently monomorphic loci could prove variable in individuals from other parts of New Zealand. The Moa_MS2 microsatellite was described previously, but had not been tested on a large sample nor assessed for allelic dropout. It was therefore included in this study to provide a full characterisation of the entire moa microsatellite marker set. The six original GS-FLX reads, from which the primers were designed, have been submitted to GenBank; the accession numbers are listed in Table 3.2.2.⁶

In an attempt to determine the chromosomal position of each of the six markers, the flanking regions were queried against the two mapped bird genomes, chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*), using BLAST (www.blast.ncbi.nlm.nih.gov/Blast.cgi). However, because of the short flanking regions and probably also because of the distant relationships of these taxa with moa, the queries were largely unsuccessful. One perhaps convincing match was that the 145 bp long flanking region in Moa_MA44 showed 83% identity with an 83 bp sequence on the zebra finch chromosome 1, and 96 % identity with a 26 bp sequence on the chicken chromosome 1. Therefore, this marker probably resides on the moa chromosome 1.

Given that the six microsatellite primer pairs worked for *Dinornis robustus*, but were designed from DNA of three different moa genera (Table 3.2.2), it was obvious that the markers possessed the ability to amplify across taxa. The primers were tested on DNA from three other moa genera (*Euryapteryx*, *Emeus* and *Pachyornis*) and amplified well. A thorough assessment of genetic diversity in the four sympatric North Canterbury moa species is underway, and comparative data will be presented elsewhere (see Chapter 5).

⁶ GenBank sequence deposition will happen when the manuscript is submitted to a journal. The six sequences can be seen in Digital Appendix 13.

Table 3.2.2: Characteristics of the six identified microsatellite markers, including the moa species in which it was identified, repeat unit as observed in the original GS-FLX clone, the annealing temperature in the 50-cycle PCR reaction from Allentoft et al. (2009), the primer sequences and the GenBank accession numbers.

Locus	Taxon	Identified as	Ann. temp.	Forward 5'→ 3'	Reverse 5'→ 3'	GenBank #
Moa_MS2	<i>P. elephantopus</i>	(AC)12	57°C	GAGCACCAATACAACCTTCATGG	GACTGTTATTCTATTCCAGTATATGTTG	XXXXXXXX
Moa_MA1	<i>D. robustus</i>	A7(CA)9	55°C	CATATAGGCACAAGGAGAGC	CAGGGGAGGATGGTATCTGT	XXXXXXXX
Moa_MA21	<i>A. didiformes</i>	(GT)9	57°C	CGTGCTCGGATGCATAGAT	GTTATCTGTGCGCCTTGC	XXXXXXXX
Moa_MA38	<i>A. didiformes</i>	(AC)8	58°C	GCTTGTTCCCTCCATCAT	CAGCACTGTGCAGCACTTTC	XXXXXXXX
Moa_MA44	<i>A. didiformes</i>	(AC)8	58°C	AGGATTAGATCCCAGGAAGC	CTCTCAGCCTGTGGACTTTG	XXXXXXXX
Moa_MA46	<i>A. didiformes</i>	(GAG)9	66°C	GGCTGTCCGCCACTCAAG	GAGAAGGGCTCGGTCCTC	XXXXXXXX

	Moa_MS2	Moa_MA1	Moa_MA21	Moa_MA38	Moa_MA44	Moa_MA46	Overall
PCR's in heterozygotes	196	175	100	203	123	100	897
PCR's with dropout	112	80	57	143	48	36	476
Observed dropout rate	0.57	0.46	0.57	0.70	0.39	0.36	0.53
PCR's with short allele dropout	39	40	29	44	23	14	189
PCR's with long allele dropout	73	40	28	99	25	22	287
Prob. of 2x same allele dropout	0.178	0.104	0.163	0.285	0.076	0.068	
Prob. of 3x same allele dropout	0.060	0.024	0.046	0.126	0.015	0.013	
Prob. of 4x same allele dropout	0.021	0.005	0.013	0.059	0.003	0.003	
Prob. of 5x same allele dropout	0.007	0.001	0.004	0.028	0.001	0.001	
Mean C _T in extracts with dropout	32.46 ± 3.08	32.89 ± 3.11	32.94 ± 3.08	32.51 ± 2.81	32.31 ± 2.81	33.10 ± 2.60	
Mean C _T in extracts without dropout	30.92 ± 2.64	31.36 ± 3.04	31.44 ± 2.07	29.75 ± 1.98	33.30 ± 3.18	31.82 ± 2.75	
<i>t</i> -tests, <i>P</i> =	0.100, df=50	0.095, df=53	0.090, df=43	0.060, df=53	0.280, df=43	0.110, df=46	0.008

Table 3.2.3: Number of PCRs of known heterozygous individuals and proportion of PCRs with dropout. Dropout rate was used to estimate the probability of misidentifying a heterozygote as a homozygote based on 2-5 PCRs. C_T mean values (± st.dev.) were calculated for DNA extracts showing *dropout* and *no dropout* respectively. *P*-values represent the significance of difference in mean values. Overall *P*-value was calculated with Fisher's approach.

3.2.4.3 Allelic dropout

In the assessment of allelic dropout, a total of 897 PCR reactions represented heterozygous individuals, according to the criteria for heterozygosity established above. Overall, 53% of the 897 PCRs showed dropout (only one allele), but the fraction varied substantially between the markers, from 36 % in Moa_MA46 to 70 % in Moa_MA38 (Table 3.2.3, Figure 3.2.1). In three of the markers (Moa_MS2, Moa_MA38, Moa_MA46), the long allele dropped out more often than the shorter one, whereas in the three other markers, the long, and short alleles dropped out with almost equal frequency (Table 3.2.3).

Given the observed dropout rates, we calculated the probability that a heterozygote would be misidentified as a homozygote because of dropout. For example, the estimated averaged probability that the same allele drops out in four consecutive PCRs when using the Moa_MS2 primers, is the probability for the long allele to drop out four consecutive times = $(73/196)^4$ plus the probability that the short allele drops out four consecutive times = $(39/196)^4$, or 2.1%, as shown in Table 3.2.3. In other words, if a 4xPCR-repeat approach was applied for the Moa_MS2 locus, a hypothetical 2.1% of the heterozygous individuals were likely to be misidentified as homozygous.

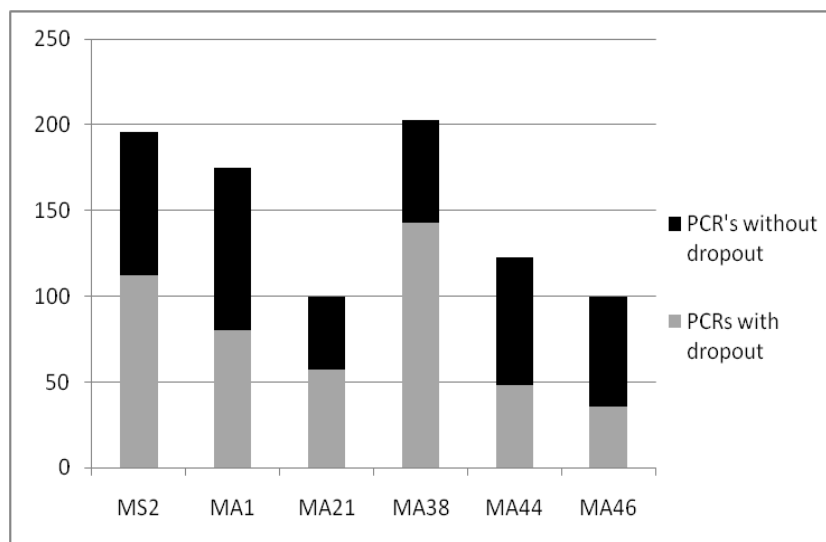


Figure 3.2.1: Graphical representation of allelic dropout in each marker. Bars show the total number of successful PCRs in heterozygous individuals, with the dropout and no-dropout fractions indicated

3.2.4.4 Establishing the criteria

With the level of allelic dropout having been assessed for each marker, there was basis for a set of rules for generating reliable moa microsatellite data. The rationale behind the three following criteria is outlined in the Discussion:

- 1) An apparent homozygote was accepted as such, only when it showed the same unambiguous mono-allelic profile in at least four independent PCR reactions.
- 2) An apparent heterozygote was accepted as such if: (i) it showed an unambiguous di-allelic profile, and/or, (ii) it displayed two different unambiguous mono-allelic profiles (representing already well-established alleles) in different PCRs.
- 3) Data from individuals that could not, despite several attempts, be reliably genotyped with at least five of the six microsatellite primers, were completely removed from the compiled dataset.

Applying these rules, we derived a multilocus microsatellite genotype for 74 of the 88 *Dinornis robustus* extracts – a success rate of 84%.

3.2.4.5 Descriptive statistics, HW-proportions and linkage

Summary statistics for the compiled data are shown in Table 3.2.4. The number of observed alleles (N_A) ranged from five in Moa_MA44 to 17 in Moa_MA38. In general, observed and expected heterozygosity proved similar, yielding only small deviations from Hardy-Weinberg proportions (F_{IS} -values). When tested, these deviations were not significant except for Moa_MA21, which displayed a small, but significant, heterozygote deficiency ($F_{IS} = 0.109$, $P = 0.04$). When Bonferroni-correction was applied to accommodate for possible statistical type I error, the deviation was no longer significant. Further, MICROCKECKER (van Oosterhout et al., 2006) found no signs of scoring errors resulting from stuttering, long allele dropout, or null alleles in the data (data not shown). Lastly, no signs of linkage disequilibrium between any of the markers were apparent, and P -values for all 15 locus combinations ranged from 0.10 to 0.96.

3.2.4.6 Quantitative PCR and dropout

To test for a relationship between DNA preservation and allelic dropout, the mtDNA CR C_T -values recorded from qPCRs were used as proxies for preservation. Of the 74 DNA extracts included in the compiled dataset, an accompanying mtDNA (Control region) C_T value was available from previous studies for 66 (Allentoft et al., 2010; Allentoft et al., 2009). Although a linkage between a high dropout rate and a high C_T value was expected, the relationship is complicated by individuals with long alleles having a higher chance of showing dropout, regardless of their preservation state. Therefore, the analysis was kept simple, with individuals separated into two groups, showing dropout

and no dropout, respectively. The average C_T values for the groups were then compared. Although not significant (t -tests) the expected pattern of a higher average C_T values in dropout-individuals was confirmed in all markers except for Moa_MA44 (Table 3.2.3, Figure 3.2.2). When P -values were combined across loci with Fisher's approach, it was demonstrated with high significance ($P = 0.008$) that individuals prone to allelic dropout were also accompanied by higher average C_T -values than individuals without observed dropout (Table 3.2.3).

Table 3.2.4: Summary statistics for six loci and 74 *Dinornis robustus* individuals. *Range* is observed allele size range, N_A is total number of observed alleles, N_E is effective number of alleles, H_O and H_E are observed and expected heterozygosity, fixation index F_{IS} measures deviations from Hardy-Weinberg proportions, with P representing the significance of this deviation. Overall P is the Fisher's P -value for combined probabilities.

Locus	n	Range	N_A	N_E	H_O	H_E	F_{IS}	P
Moa_MS2	72	110-148 bp	16	4.8	0.778	0.794	0.020	0.674
Moa_MA1	74	91-99 bp	7	3.5	0.730	0.715	-0.020	0.897
Moa_MA21	74	93-113 bp	6	3.3	0.622	0.698	0.109	0.040
Moa_MA38	72	84-132 bp	17	6.1	0.792	0.837	0.054	0.331
Moa_MA44	74	75-85 bp	5	2.8	0.595	0.641	0.073	0.257
Moa_MA46	74	66-81 bp	6	2.6	0.608	0.610	0.003	0.238
Overall	74	66-148 bp	9.5	3.8	0.687	0.721	0.040	0.227

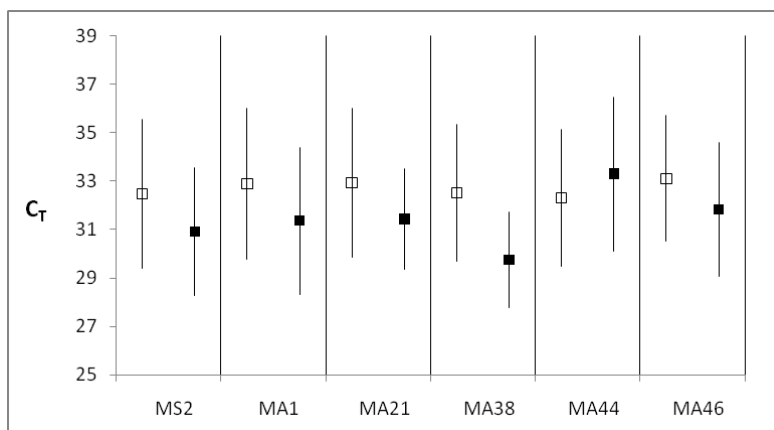


Figure 3.2.2: Graphical representation of differences in average C_T values (\pm st. dev.) in extracts with allelic dropout (empty boxes) and extracts with no dropout (black boxes). Higher C_T values are accompanied by higher chance of dropout in five of the six markers. Exact values are listed in Table 3.2.3

This confirmation allowed for a direct investigation of the integrity of the full *Dinornis* dataset. If high individual homozygosity (proportion of homozygote loci in an individual) was an effect of dropout, a negative correlation would be expected between observed homozygosity and the recorded

C_T value. The results presented in Figure 3.2.3 show that this is not so ($R^2 = 0.017$). DNA preservation in highly homozygous individuals was no worse than in highly heterozygous individuals.

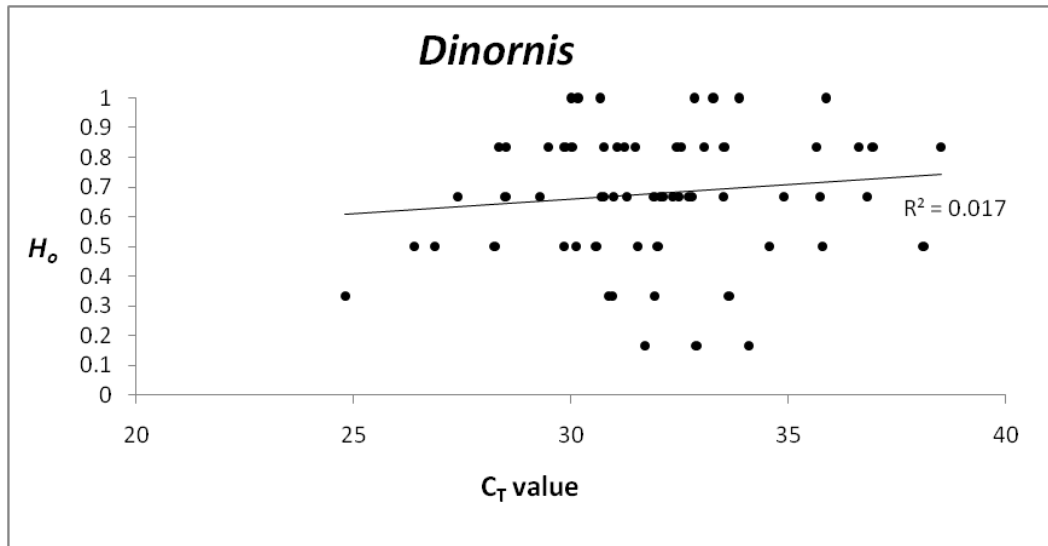


Figure 3.3.3: Evidence that observed individual heterozygosity (H_o) among 66 *Dinornis* individuals in the final dataset is not linked to DNA preservation (represented by C_T values). If that was the case, a clear negative correlation would be expected here.

3.2.5 Discussion

From this study, we have shown that it is possible to generate a fully functional microsatellite library from aDNA. Several studies have amplified microsatellite markers from museum specimens, for example to investigate a loss of genetic biodiversity resulting from recent environmental changes (e.g., Bourke et al., 2010; Pertoldi et al., 2001; Petren et al., 2010; Whitehouse, Harley, 2001). Here we show that reliable microsatellite data can be generated from samples of much greater antiquity, but that considerable care needs to be exercised in compiling the data.

3.2.5.1 High-throughput sequencing

This study clearly demonstrates that it takes a significant effort to retrieve appropriate microsatellite markers from ancient DNA. From >600,000 GS-FLX sequences, >1800 detected STRs, and 89 tested primer-pairs, a total of six functional microsatellite markers may seem a rather poor return. This difficulty is largely a consequence of the fragmented nature of ancient DNA, but also reflects the

capability of the GS-FLX Standard platform in terms of read lengths. The more recent GS-FLX Titanium chemistry (454 Life Sciences of Roche, Branford, CT, USA), with longer read lengths, is likely to improve the success rate. Differing DNA quality probably explains the observed variation in results between extracts in Table 3.2.1. In terms of number of reads, average read lengths, and number of detected STRs, the eggshell extracts appear to perform better than the bone extracts, apart from one very poor eggshell extract, with a short average read length (MB 595). This agrees with a recent study demonstrating high DNA quality in ancient eggshells, including moa (Oskam et al., 2010). The *Anomalopteryx* eggshell (MB 718) produced many reads, and yielded four of the six microsatellites described here.

Whereas our combined data showed an average of 0.3% of the sequences containing repeats (1809 of 603773 sequences), values such as 2.4% (Duckett, Stow, 2010), 1.8% (Clay et al., 2010), 1.3% (Abdelkrim et al., 2009), and 11.3% (Castoe et al., 2009) have been demonstrated when mining for microsatellites in modern DNA. Although the numbers are not strictly comparable because of slight variations in the STR search-criteria, it seems that ancient DNA produces a considerably smaller fraction of sequences with STRs. Microsatellites are found only in the nuclear genome, so this difference could perhaps reflect the scarcity of nuclear DNA in the fossils, or that microbial DNA (in which STRs are rare – see Chapter 3, Section 1) contributes to a high proportion of the reads, or perhaps that microsatellites were simply relatively rare in moa, as has been suggested for other birds (Primmer et al., 1997).

3.2.5.2 Allelic dropout

It is clear, from the data in Table 3.2.3, that allele scoring was seriously hampered by allelic dropout. Indeed, if the common approach with just one PCR reaction per marker was applied, an average of 53% of the heterozygotes would likely have been misidentified as homozygotes. Likewise, attempts to multiplex just two of these markers together caused even greater dropout. As mentioned in the Introduction, the problem is well described in the literature although the average dropout we observed seems to have been slightly worse than in studies of, for example, faecal DNA [24% dropout in Morin et al. (2001) and 29% dropout in Frantz et al. (2003)], DNA from 100-year-old teeth [42% dropout in Arandjelovic et al. (2009)], shed hair [31% dropout in Gagneux et al. (1997)] and human fingerprints [12% in Balogh et al. (2003)]. The three criteria, proposed to minimise the effect of dropout on our moa data are discussed below:

Criterion 1 involved a 4x singleplex PCR repeat method (multiple tubes approach) for each apparent homozygote. According to the average error rates shown in Table 3.3.3, this should reduce

genotyping errors to between 0.3% (Moa_MA 44 and Moa_MA46) and 5.9% (Moa_MA21) in the generated data. This 4x approach was based on a balance between generating data of high quality, while maintaining enough DNA extract to amplify all six microsatellites markers. The evaluation had to factor in the relative difficulty in obtaining sample material for additional extractions, and also account for a high rate of PCR failures, given that retrieving results from four positive PCR reactions could for some extracts easily involve 6-10 PCR setups. Our approach here is less conservative than the often cited 7x PCR repeat suggested by Taberlet et al. (1996), which was based on computer simulations of a worst-case scenario with 100 % allelic dropout. Rather than automatically assuming a worst-case scenario, we measured the level of dropout directly in each marker and relied on a less stringent multiple tubes approach, more comparable to protocols suggested elsewhere (e.g., Frantz et al., 2003; Miller et al., 2002).

Criterion 2 included the genotyping of heterozygotes, which we accepted as individuals showing two distinct (and already well-established) alleles in any number of PCR reactions (i.e. two different alleles in the same PCR or two different alleles in separate PCRs). Other studies have suggested stricter rules, arguing, for example, that an allele must be recorded at least twice to be accepted (Frantz et al., 2003; Taberlet et al., 1996). For a large proportion of our samples, each allele was indeed recorded more than once, but we found no support in the data for applying this as a criterion. The concern in this context is that a contamination event or lack of specificity in the PCR reaction can lead to scoring of false alleles. Both these problems might theoretically arise more frequently in PCRs using aDNA templates. However, the problem appeared so rarely that it should not set a precedent for a rule. Amongst the >1500 successful PCR reactions, we recorded less than 10 reactions with an odd allele that could not be accounted for easily. *False* peaks appeared regularly though with two of the markers (such as a distinct 115 bp peak with Moa_MA38), but these lacked the characteristics of microsatellite profiles, such as stutter peaks, and could hence easily be identified and eliminated. Contamination is potentially a more serious problem and it is possible that an allele observed only once represented contamination from another sample. However, this also applies to an allele observed twice. If a sample is contaminated, multiple PCRs will not necessarily clarify the situation. In contrast, if cross-contamination happens occasionally during PCR setup, confirmation in multiple repeats should effectively remove its effect. We are, however, confident that there was very little cross-contamination evident in our data. Because of the difficulty in obtaining moa samples, DNA extracts were stored in separate 1.5 ml Eppendorf tubes to minimise evaporation. Each tube was opened individually and DNA extract was single-pipetted into each respective PCR, which involved much less risk of cross-contamination than do the commonly applied multi-pipette PCR setups. Additionally, extractions and PCRs were performed in a dedicated clean lab using strict aDNA guidelines. These strict measures may explain why we did not observe contamination issues in the

Dinornis data, either here nor in previous work based on the same DNA extracts (Allentoft et al., 2010; Allentoft et al., 2009).

Criterion 3 involved the removal of problematic individuals from the dataset. Rather than using indirect indications for the potential genotypic reliability of each sample, based, for example, on the recorded mtDNA CR qPCR C_T values, we took a more direct approach, and removed individuals that could not amplify in at least five of the six microsatellites, despite several attempts. It is clear that successful PCR amplifications in different loci constitute strong reciprocal support for the results. For example, if a 144 bp allele in the *Moa_MS2* locus has been identified in a particular extract, it is a confirmation that sufficient nuclear DNA preservation exists for long alleles in the other loci to be amplified as well. On the other hand, if a given DNA extract can amplify only one or two of the six microsatellites despite repeated attempts, the argument collapses and data for that individual should be removed from the compiled dataset. Differences in primer efficacy and in the presence/absence of PCR inhibitors in some ancient DNA extracts can exacerbate these problems.

3.2.5.3 The data

A critical evaluation of the final data confirmed its integrity. One minor deviation from Hardy-Weinberg proportions was observed in the *MA21* locus, but as these data refer to birds living over a span of >4000 years, they do not reflect randomly mating individuals at one point in time; so HW-proportions are not necessarily expected by default. Yet, the general accordance with expected HW-proportions was an indication that the impact of allelic dropout was minimal. Use of additional software, designed specifically to identify scoring problems because of dropout and stuttering, did not reveal any of these issues. Lastly, we documented a link between dropout and C_T values, but clearly rejected a negative correlation between observed individual homozygosity and C_T . This also provided strong support for the integrity of the data.

We suggest therefore that the *moa* microsatellite dataset is of high quality despite representing template molecules of app. 600 to 5000 years of age (see Chapter 4). We argue that variants of the applied methodology will be valid for most scenarios involving aDNA and microsatellites. We emphasise, however, the importance of assessing each case on its merits, including pilot studies and the generation of preliminary data to develop a specific strategy for the material at hand. Criterion 2 for example, is likely to be inappropriate for data where false peaks are more difficult to discriminate from true alleles or where contamination is of greater concern.

The achievement of generating a high quality microsatellite dataset for an extinct species, does not mean that there is no room for methodological or procedural improvement. A major drawback to the

procedure presented here, is the considerable workload associated with single-plexed PCRs, repeated many times with DNA added from one tube at a time. A few experiments with two-plexes were, however, unsuccessful. Methods have been developed for improving PCR results from degraded DNA (e.g., Arandjelovic et al., 2009; Roeder et al., 2009); further research might show whether our stringent setup could be relaxed somewhat, and whether these novel approaches can be used on aDNA templates without increasing the risk of cross-contamination. However, to generate this first aDNA microsatellite data set, we applied a very simple framework to minimise the number of unknown factors that could compromise the novel results.

3.2.6 Conclusion

With the six markers presented here, the tools are now available for conducting a series of high resolution genetic studies of the New Zealand megafauna. We are aware that a set of six polymorphic markers are smaller than commonly applied in present day population genetics but are still unlike anything previously attempted in aDNA research. Secondly, more potential markers have been identified and may, in time, become available for research.

Further analyses with these data represent another challenge. Most available analytical methods rely on allele frequencies representing random union of gametes, and this is certainly not so when data are from individuals spanning hundreds of generations. A new viewpoint will be required to take advantage of the information in heterochronous microsatellite data, for example by applying approximate Bayesian computation (ABC) (e.g., Anderson et al., 2005).

Although the moa fossils included here are of late Holocene age, and therefore younger than fossils from all the extinct Pleistocene megafauna species, moa were not preserved in very cold environments such as those typical of, for example, woolly mammoth (*Mammuthus primigenius*), woolly rhino (*Coelodonta antiquitatis*) or steppe bison (*Bison priscus*). With the significant success in amplifying nuclear microsatellite DNA we document here, and knowing that low temperatures greatly increases the longevity of ancient DNA, it seems very likely that our technology can be applied to research on many other extinct taxa, if they have been preserved under the right conditions.

3.2.7 References

- Abdelkrim J, Robertson BC, Stanton JAL, Gemmell NJ (2009) Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *Biotechniques* **46**, 185-192.
- Allentoft ME, Bunce M, Scofield RP, Hale ML, Holdaway RN (2010) Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand's extinct megafauna. *Quaternary Science Reviews* **29**, 753-762.
- Allentoft ME, Schuster S, Holdaway RN, et al. (2009) Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *Biotechniques* **46**, 195-200.
- Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA (2005) Serial SimCoal: A population genetics model for data from multiple populations and points in time. *Bioinformatics* **21**, 1733-1734.
- Arandjelovic M, Guschanski K, Schubert G, et al. (2009) Two-step multiplex polymerase chain reaction improves the speed and accuracy of genotyping using DNA from noninvasive and museum samples. *Molecular Ecology Resources* **9**, 28-36.
- Balogh MK, Burger J, Bender M, Schneider PM, Alt KW (2003) STR genotyping and mtDNA sequencing of latent fingerprint on paper. *Forensic Science International* **137**, 188-195.
- Bourke B, Frantz A, Lavers C, et al. (2010) Genetic signatures of population change in the British golden eagle (*Aquila chrysaetos*). *Conservation Genetics* DOI 10.1007/s10592-010-0076-x.
- Brennan PLR, Hyseni C (2008) Development of microsatellite markers for parentage analysis in the great tinamou (*Tinamus major*). *Molecular Ecology Resources* **8**, 933-935.
- Bunce M, Worthy TH, Phillips MJ, et al. (2009) The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20646-20651.
- Castoe TA, Poole AW, Gu WJ, et al. (2009) Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. *Molecular Ecology Resources* **10**, 341-347.
- Clay C, Gleeson D, Howitt R, et al. (2010) Characterisation of microsatellite markers for the primitive New Zealand frog, *Leiopelma hochstetteri*. *Conservation Genetics Resources* DOI 10.1007/s12686-010-9211-8.
- Dawson DA, Horsburgh GJ, Kupper C, et al. (2010) New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utility - as demonstrated for birds. *Molecular Ecology Resources* **10**, 475-494.
- Duckett P, Stow A (2010) Rapid isolation and characterisation of microsatellite loci from a widespread Australian gecko, the Tree Dtella, *Gehyra variegata*. *Conservation Genetics Resources* DOI 10.1007/s12686-010-9248-8.
- Faircloth BC (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* **8**, 92-94.
- Frantz AC, Pope LC, Carpenter PJ, et al. (2003) Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Molecular Ecology* **12**, 1649-1661.
- Gagneux P, Boesch C, Woodruff DS (1997) Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology* **6**, 861-868.
- Goossens B, Waits LP, Taberlet P (1998) Plucked hair samples as a source of DNA: reliability of dinucleotide microsatellite genotyping. *Molecular Ecology* **7**, 1237-1241.
- Holdaway RN, Jacomb C (2000) Rapid extinction of the moas (Aves : Dinorinithiformes): Model, test, and implications. *Science* **287**, 2250-2254.

- Huang Y, Liu Q, Tang B, et al. (2008) A preliminary microsatellite genetic map of the ostrich (*Struthio camelus*). *Cytogenetic and Genome Research* **121**, 130-136.
- Kimwele CN, Graves JA, Burke T, Hanotte O (1998) Development of microsatellite markers for parentage typing of chicks in the ostrich *Struthio camelus*. *Molecular Ecology* **7**, 249-251.
- Maricic T, Paabo S (2009) Optimization of 454 sequencing library preparation from small amounts of DNA permits sequence determination of both DNA strands. *Biotechniques* **46**, 51-+.
- Meyer M, Briggs AW, Maricic T, et al. (2008) From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Research* **36**.
- Miller CR, Joyce P, Waits LP (2002) Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics* **160**, 357-366.
- Morin PA, Chambers KE, Boesch C, Vigilant L (2001) Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology* **10**, 1835-1844.
- Navidi W, Arnheim N, Waterman MS (1992) A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR - statistical considerations. *American Journal of Human Genetics* **50**, 347-359.
- Oskam CL, Haile J, McLay E, et al. (2010) Fossil avian eggshell preserves ancient DNA. *Proc Biol Sci* **277**, 1991-2000.
- Parsons KM (2001) Reliable microsatellite genotyping of dolphin DNA from faeces. *Molecular Ecology Notes* **1**, 341-344.
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.
- Pertoldi C, Hansen MM, Loeschcke V, et al. (2001) Genetic consequences of population decline in the European otter (*Lutra lutra*): an assessment of microsatellite DNA variation in Danish otters from 1883 to 1993. *Proceedings of the Royal Society B-Biological Sciences* **268**, 1775-1781.
- Petren K, Grant PR, Grant BR, Clack AA, Lescano NV (2010) Multilocus genotypes from Charles Darwin's finches: biodiversity lost since the voyage of the Beagle. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**, 1009-1018.
- Primmer CR, Raudsepp T, Chowdhary BP, Moller AR, Ellegren H (1997) Low frequency of microsatellites in the avian genome. *Genome Research* **7**, 471-482.
- Roeder AD, Elsmore P, Greenhalgh M, McDonald A (2009) Maximizing DNA profiling success from sub-optimal quantities of DNA: A staged approach. *Forensic Science International-Genetics* **3**, 128-137.
- Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.
- Santana QC, Coetzee MPA, Steenkamp ET, et al. (2009) Microsatellite discovery by deep sequencing of enriched genomic libraries. *Biotechniques* **46**, 217-223.
- Sefc KM, Payne RB, Sorenson MD (2003) Microsatellite amplification from museum feather samples: Effects of fragment size and template concentration on genotyping errors. *Auk* **120**, 982-989.
- Taberlet P, Griffin S, Goossens B, et al. (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**, 3189-3194.
- Tang B, Huang YH, Lin L, et al. (2003) Isolation and characterization of 70 novel microsatellite markers from ostrich (*Struthio camelus*) genome. *Genome* **46**, 833-840.
- Taylor EL, Vercoe P, Cockrem J, et al. (1999) Isolation and characterization of microsatellite loci in the emu, *Dromaius novaehollandiae*, and cross-species amplification within Ratitae. *Molecular Ecology* **8**, 1963-1964.

- Untergasser A, Nijveen H, Rao X, et al. (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research* **35**, W71-W74.
- van Oosterhout C, Weetman D, Hutchinson WF (2006) Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Molecular Ecology Notes* **6**, 255-256.
- Ward WK, McPartlan HC, Matthews ME, Robinson NA (1998) Ostrich microsatellite polymorphisms at the VIAS-OS4, VIAS-OS8, VIAS-OS14, VIAS-OS22, and VIAS-OS29 loci. *Animal Genetics* **29**, 331-331.
- Whitehouse AM, Harley EH (2001) Post-bottleneck genetic diversity of elephant populations in South Africa, revealed using microsatellite analysis. *Molecular Ecology* **10**, 2139-2149.
- Wood JR, Worthy TH, Rawlence NJ, Jones SM, Read SE (2008) A deposition mechanism for Holocene miring bone deposits, South Island, New Zealand. *Journal of Taphonomy* **6**, 1-20.
- Worthy TH, Holdaway RN (1996) Quaternary fossil faunas, overlapping taphonomies, and palaeofaunal reconstruction in north Canterbury, South Island, New Zealand. *Journal of the Royal Society of New Zealand* **26**, 275-361.
- Worthy TH, Holdaway RN (2002) *The lost world of the Moa* Canterbury University Press, Christchurch.

Chapter 4: Mitochondrial DNA analyses of four sympatric moa populations

Mitochondrial DNA analyses of four sympatric moa populations

4.1 Introduction

Chapters 4 and 5 in this thesis contain the first studies of moa population genetics. Although the results in Chapter 2 were interpreted in the context of populations, they were based on simple molecular identifications of species and gender, without using the information from the genetic diversity *per se*. Here, and in Chapter 5, the genetic variability within each of the four sympatric moa populations will be examined at several levels. Rather than combining all the population genetics in a single manuscript, the mtDNA analyses and microsatellite analyses have been separated. This was done to allow for a thorough, 'basic' exploratory investigation, identifying the strengths, potentials, and caveats in these unique data, while at the same time exposing areas where more exhaustive analyses are required.

There has been a strong preference for targeting mitochondrial DNA (mtDNA) in molecular research based on ancient substrates, because multiple mitochondrial genome copies are present in each cell. It is therefore less challenging to amplify than nuclear DNA (reviewed in Hofreiter et al., 2001; Willerslev, Cooper, 2005). Mitochondrial genome regions such as the *cytochrome b* (*cytb*) gene or the hypervariable non-coding Control region (CR) have been genetic markers of choice in many aDNA studies, for example to establish the taxonomic relationships between extinct and extant species (e.g., Bunce et al., 2005; Seabrook-Davison et al., 2009; Shapiro et al., 2002), to infer intraspecific phylogeographic patterns (e.g., Barnett et al., 2009; Bunce et al., 2009; Debruyne et al., 2008; Lambert et al., 2002), or to assess demographic histories of ancient populations (e.g., de Bruyn et al., 2009; Shapiro et al., 2004; Stiller et al., 2010). Prior to this thesis, publications on moa genetics have relied almost exclusively on mtDNA (Table 1.1.1), except for two contributions applying molecular sexing (Bunce et al., 2003; Huynen et al., 2003). Because the mt-genome is haploid and maternally inherited, the effective population size of a mitochondrial locus is effectively $\frac{1}{4}$ of a nuclear one (Hedrick, 2000). Mutations therefore tend to accumulate more rapidly, making mtDNA sequences highly informative, even in population-level contexts (Ho, Gilbert, 2010). However, a maternally-inherited molecule such as the mt-genome will only reflect the females in the population and gene flow facilitated by males will not be revealed. Moreover, the haploid and non-recombining nature of mtDNA means that old and genetically very different maternal lineages can be present in a population with otherwise perfect genetic admixture. This is evident in, for example, the woolly mammoth (*Mammuthus primigenius*) gene pool which shows highly divergent mtDNA lineages that were sympatric in space and time (Gilbert et al., 2008). Interpretations regarding population structure and dispersal events must therefore be cautious when based entirely on mtDNA analyses.

In this chapter, the ^{14}C -radiocarbon data are presented first, providing a temporal framework for all the remaining analyses. The concordance between the timing of moa deposition in these natural North Canterbury sites and the arrival of humans in New Zealand will be discussed. This is relevant to address the timing of local extinction. Following that, several aspects of moa population genetics will be assessed:

First, the genetic diversity in the control region sequences will be treated in an interspecific comparative analysis. It has been suggested that the Holocene population of *Emeus crassus* showed less genetic variability than other moa species, but this was based on a very small sample size (Bunce et al., 2009). However, if indeed this species displayed disproportionately low genetic diversity, it would suggest that it went through a bottleneck some time prior to the late Holocene, and further; that whatever environmental factors caused such decline, it affected this species more than the others. To address this, genetic diversity in the four North Canterbury populations will be quantified and compared based on both mtDNA (this chapter) and microsatellite loci (Chapter 5).

Next, the genetic structure of the four populations will be investigated to see if the structure was defined by either space or time. This is examined with mtDNA haplotype networks by combining the DNA sequence information with information on the origin (site) and age (^{14}C data) from each individual. As null hypothesis, the genetic diversity is not expected to have been influenced by geography, because the fossil sites are so close together and their hinterlands were very unlikely supporting distinct “colonies” of moa without connecting gene flow. Rather, they represent independent windows into the structure of the same surrounding prehistoric moa assemblage. If any genetic population structure is present in the data, a time-related structuring seems more plausible. Climatic or other environmental changes could have facilitated sudden shifts in genetic makeup, resulting, for example, from a temporary local absence of moa, followed by re-colonisations of new genetic variants. Such distinct temporal replacements of mitochondrial lineages have been described in cichlid fish after a long time of separation followed by reunion and introgression (Nevado et al., 2009), in a continuous population of mice adapting to anthropogenically altered habitats (Pergams et al., 2003), in Pleistocene cave bears, probably following disappearance and re-colonisation in response to human activity (Hofreiter et al., 2007), and even in early European human populations, showing genetic discontinuity between Paleolithic hunter-gatherers and the populations who replaced them (Bramanti et al., 2009).

The data will then be analysed for a demographic signal by applying a Bayesian, coalescent-based, method, which has previously proven successful in revealing past demographics from ancient DNA, (e.g., Shapiro et al., 2004; Stiller et al., 2010). Because, the moa extinction was very rapid (Holdaway, Jacomb, 2000), it was not expected that the process left genetic traces accessible to these analyses. Moa probably had long generation times (Turvey et al., 2005; Turvey, Holdaway, 2005), requiring

decades of persistent (but not exterminating) hunting pressure on the local populations, for genetic drift to manifest itself as lowered genetic diversity. Additionally, the brief moa-human interaction phase is likely to be represented by very few fossils in non-archaeological contexts. So, rather than profiling the short, intense extinction process, the demographics are more likely to reflect moa populations in the pristine Holocene environment of New Zealand, before the arrival of Polynesians. However, as mentioned previously, Gemmell et al. (2004) argued that the moa populations collapsed suddenly in the late Holocene, before humans invaded the ecosystem. This hypothesis of a pre-human decline will be assessed here (and again in Chapter 5), on the basis of all the North Canterbury moa data. The datasets will be thoroughly examined for any traces of a massive pre-human, late Holocene population collapse. Moreover, an attempt to estimate the population size of the South Island giant moa (*Dinornis robustus*) will be carried out.

Finally, the sequence data for each of the four moa species in North Canterbury will be analysed in a wider geographic context by including all the relevant moa CR sequences available from GenBank. If any geographic structuring is present on the South Island, it will be described and interpreted briefly in this chapter. Importantly, moa phylogeography is not among the main foci in this thesis. Indeed, the strength and novelty of these genetic data is that they represent a small geographic area, allowing for changes in genetic variability to be monitored in only one dimension, namely time. However, there are still good reasons for extending the research beyond North Canterbury. For example, it is relevant to know whether an estimate of effective population size reflects a single panmictic population covering the entire South Island, or only a subpopulation. Also, if South Island panmixia can be refuted, it will question the theoretical framework on which Gemmell et al. (2004) based their population size estimates (see discussion in Section 4.3.4). Two previous studies have already touched on moa phylogeography, but they focused mainly on the timing and causes of radiation rather than on intraspecific signals (Baker et al., 2005; Bunce et al., 2009). With the data generated on North Canterbury moa in present project, the number of available moa control region sequences has effectively been doubled, providing a more powerful dataset than has been available before, for investigating these patterns.

4.2 Methods and results

4.2.1 ^{14}C chronology

Of 290 sampled moa bones, 217 were ^{14}C AMS dated, using protocols set out in Chapter 2, (Section 2). Radiocarbon age data associated with each individual bone are found in Digital Appendix 2. Table 4.1 provides an overview of the results, covering a calibrated chronology of 12364 years from 12966 BP to 602 BP. Including the 95% probability limits on the oldest and youngest dates respectively, the

time span is 12525 calendar years. However, as seen in Figure 4.1, only 14 of the 217 dated fossils were older than 4000 BP, with the two oldest dated to 12966 BP (Glencrieff) and 7839 BP (Rosslea) respectively. The Glencrieff bone was the only Pleistocene fossil sampled for this study. Most fossils (90 %) sampled here were younger than 3400 years BP, and hence of late Holocene origin. Figures 4.1a and b show the age data by fossil site and taxon respectively. Despite the limited number of dates from each, it is clear that Rosslea and Glencrieff contain the oldest material. The Glenmark deposit seems to cover the entire period from mid-Holocene to moa extinction. The two main sites, PV and BHV appear to have “sampled” moa slightly asynchronously, contributing with large fractions of fossils at different times (Figure 4.1a). Table 4.1b shows that all species are present throughout the Holocene, confirming the expected sympatry.

Except for one previously dated Glenmark bone, 7110 ± 110 BP (see Chapter 1, Section 3), all previous radiocarbon-dates from the five sites fell within the timeframes defined by the new data,

Table 4.1a: Radiocarbon-ages from each site. The median calibrated ages (OxCal ver.4.1) are shown with 95% probability distributions for the youngest and oldest fossil respectively. Data generated in this study (see Chapter 2, Section 2 for method).

Site	# sampled	# dated	Age	Age 95%
Pyramid Valley	147	127	602-4136 BP	545-4239 BP
Bell Hill Vineyard	121	74	851-2876 BP	796-2967 BP
Rosslea	13	8	1482-7839 BP	930-7938 BP
Glenmark	7	6	782-5152 BP	733-5312 BP
Glencrieff	2	2	4837-12966 BP	4712-13070 BP
Total	290	217	602-12966 BP	545-13070 BP

Table 4.1b: Radiocarbon-ages per taxon. The median calibrated (OxCal ver.4.1) ages are shown with 95% probability distributions for the youngest and oldest fossils respectively. Data generated in this study (see Chapter 2, Section 2 for method).

Taxon	# sampled	# dated	Age	Age 95%
<i>D. robustus</i>	89	88	602-12966 BP	545-13070 BP
<i>P. elephantopus</i>	32	32	853-4067 BP	791-4227 BP
<i>E. curtus</i>	84	40	945-5812 BP	804-5915 BP
<i>E. crassus</i>	85	57	874-5791 BP	745-5987 BP
Total	290	217	602-12966 BP	545-13070 BP

Figure 4.1a

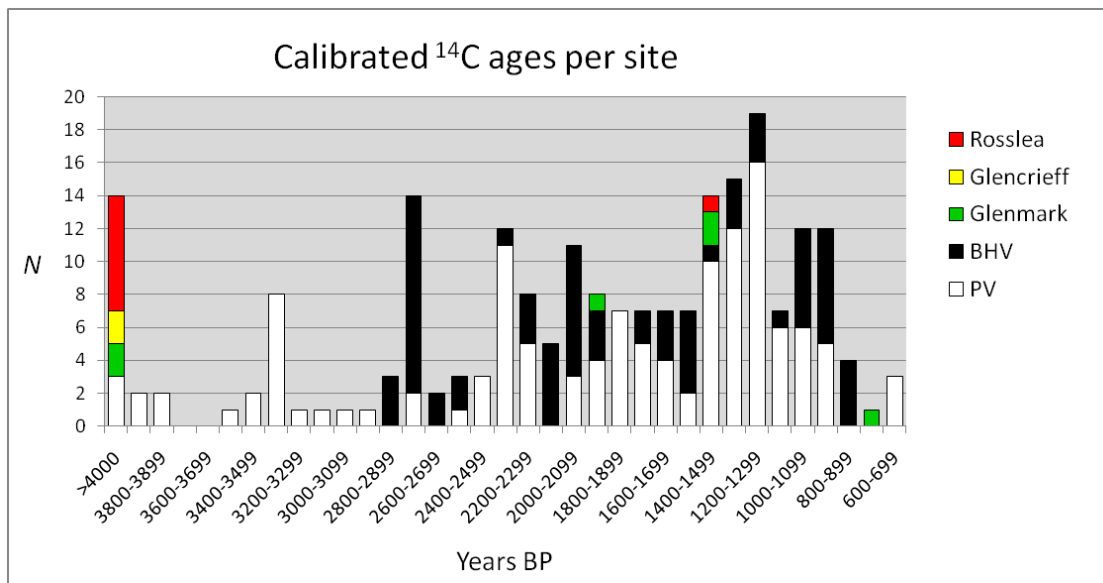
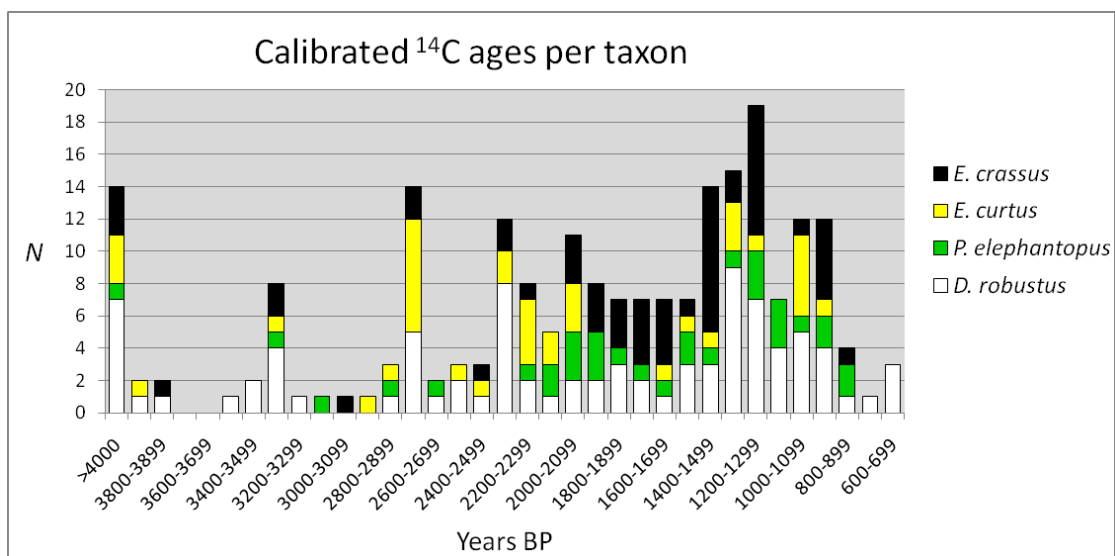


Figure 4.1b



Figures 4.1a,b: Distribution of median calibrated radiocarbon ages (a) from each site and (b) per taxon. Although representing the same overall timeframe (BHV slightly younger), PV and BHV appear to have alternated in moa deposition. All four taxa are present at all times, confirming the expected sympatry throughout the Holocene.

4.2.2 Genetic diversity

The 340 bp mtDNA CR fragment was amplified using a combination of moa specific primers (185F, 262F, 294R, 441R) with PCR conditions described previously (Bunce et al., 2003; Cooper et al., 2001). All procedures for bone sampling, DNA extraction, PCR, cloning, sequencing and species identification have been established and described in Chapter 2.

Of 290 sampled individuals, the entire targeted control region fragment was amplified from 281 (96.8 %) (Table 4.1). DNA sequences with ambiguous bases that could not be resolved by repeated PCRs or cloning, were excluded from further analyses. Moreover, mutations observed fewer than three times in each taxon were only accepted if recorded in at least one additional PCR reaction. This protocol largely removed the possibility of including false genetic diversity in the form of miscoding lesions or sequencing errors (see Chapter 2). A proportion of c. 20 % of these rare polymorphisms could not be confirmed in subsequent PCRs, which showed a common nucleotide instead. The 281 sequences were aligned for each species separately using default settings in GENEIOUS ver. 4.8.3 (Drummond et al., 2010) and the alignments were inspected manually and trimmed. Sequence alignments are given in Digital Appendix 1. The alignment files were then imported into DNASP ver. 5.10 (Librado, Rozas, 2009) and measures of genetic diversity were calculated (Table 4.2). The number of haplotypes (h) within each species ranged from 29 in *Dinornis* to 11 in *Emeus*, (0.33 and 0.14 respectively, when referenced to sample size, h/n). The number of segregating sites (S), ranged from 24 in *Dinornis* to 9 in *Emeus*. Ten and one of these respectively were only present in a single sequence (singletons). Values of nucleotide diversity (π), representing the average number of nucleotide differences per site between two sequences, and the average number of nucleotide differences between two sequences (k) showed that *Dinornis* displayed the highest genetic diversity and *Emeus* less diversity than the other three species.

Table 4.2: Measures of mtDNA (control region) genetic diversity in the four sympatric North Canterbury moa populations. The number of complete and unambiguous sequences (n) is shown with the number of observed haplotypes (h) and h/n is shown for interspecific comparative purposes. S is the number of segregating sites, with *singletons* only observed once. The nucleotide diversity (π) and the average number of nucleotide differences between two sequences (k).

Taxon	sampled	n	length	h	h/n	S	singletons	π	k
<i>D. robustus</i>	89	87	341 bp	29	0.33	24	10	0.010	3.44
<i>P. elephantopus</i>	32	31	337 bp	13	0.42	15	4	0.008	2.55
<i>E. curtus</i>	84	82	338 bp	16	0.20	15	7	0.007	2.29
<i>E. crassus</i>	85	81	337 bp	11	0.14	9	1	0.004	1.19

To visualise the mitochondrial genetic diversity in the four North Canterbury moa populations, median-joining haplotype networks were generated by importing and analysing the alignments in NETWORK ver. 4.5 and modified in NETWORK PUBLISHER ver. 1.2 (<http://www.fluxus-engineering.com>). NETWORK reconstructs the shortest and least complex evolutionary history by employing a maximum parsimony (MP) based algorithm (Figure 4.2).

In general, the data reflected good coverage of the gene pools with most intermediate haplotypes present. Signs of genetic structuring were not obvious except for *Dinornis robustus*, which showed two clades (A and B) of distinct lineages separated by 3-4 mutations (depending on the evolutionary pathway), and a single genetically distinct individual (C) perhaps representing a separate haplogroup (Figure 4.2a). Also, one Rosslea *Euryapteryx curtus* individual was very distinct, separated by five unique mutations (Figure 4.2d). To investigate whether the genetic structure was influenced by time or space, colour coding was used to visualise the associations between haplotype, radiocarbon age, and provenance. The networks including radiocarbon age and locality information are shown for *D. robustus* (Figures 4.2a and b), whereas only age-coding is shown for the emeids (*Pachyornis*, *Euryapteryx*, *Emeus*) (Figures 4.2c-e). Age groups and provenance appeared to be distributed randomly across the haplotype networks, implying that these features were not influencing the genetic structure. In *Dinornis*, for example, the most common haplotype was present in all five radiocarbon age groups, and four of the five sites (Figures 4.2a and b). The difference in genetic diversity between taxa was clearly reflected in the networks as well, notably by the very low variability in the *Emeus crassus* gene pool, which is dominated by two common haplotypes separated by a single mutation (Figure 4.2e).

A star-like pattern was observed in mitochondrial lineage A in the *Dinornis* network (Figures 4.2a and b). A single common haplotype is associated with many rare satellite haplotypes, separated by just 1-2 mutations from the common one. Theoretical predictions (e.g., Slatkin, Hudson, 1991), as well as empirical studies (e.g., Mirol et al., 2008), have previously suggested such star-like clusters around a founder node to be the resulting of a population expansion (see Section 4.3.4).

Figure 4.2a: *Dinornis robustus*, age

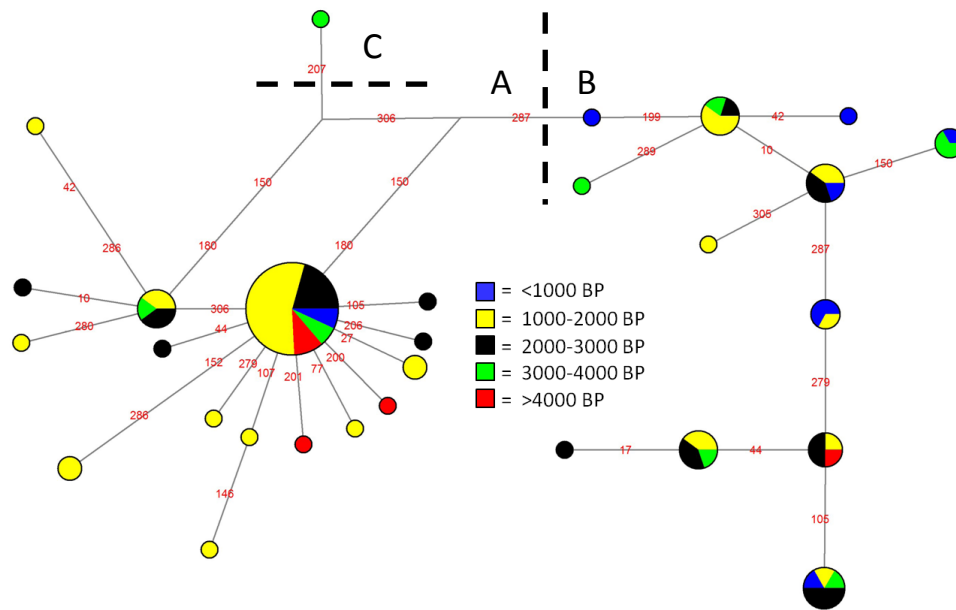
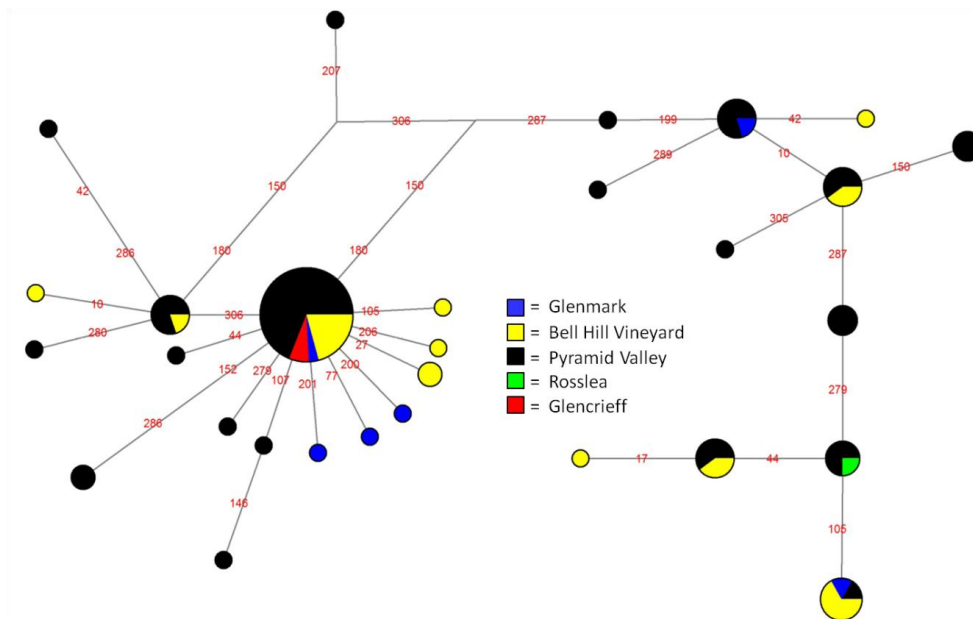
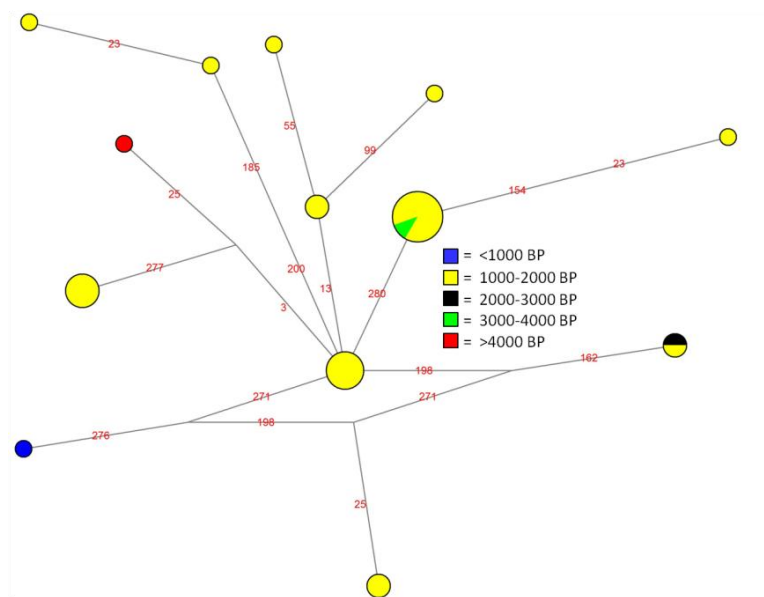


Figure 4.2b: *Dinornis robustus*, provenance



Figures 4.2a,b: Median-joining haplotype-networks based on 87 North Canterbury individuals of the South Island giant moa (*Dinornis robustus*). Twenty nine distinct haplotypes were present with size of pies determined by the number of individuals sharing each haplotype. Each individual has been colour coded according to the calibrated ^{14}C age (a) or according to the provenance of the associated fossil (b). Numbers on the links is positions of mutations in the DNA sequences, separating the haplotypes. Note the separation into two main haplogroups (lineages A and B), a single individual from a rare haplogroup (C), and a star-like pattern in lineage A.

Figure 4.2c: *Pachyornis elephantopus*



Figures 4.2c,d,e: Median-joining haplotype networks of the emeids; *Pachyornis elephantopus* (3c), *Euryapteryx curtus* (3d) and *Emeus crassus* (3e). Colour coding according to calibrated ^{14}C age with white areas representing sequences without associated ages. Note a single distinct *Euryapteryx* individual, and the low genetic diversity in the *Emeus* population.

Figure 4.2d: *Euryapteryx curtus*

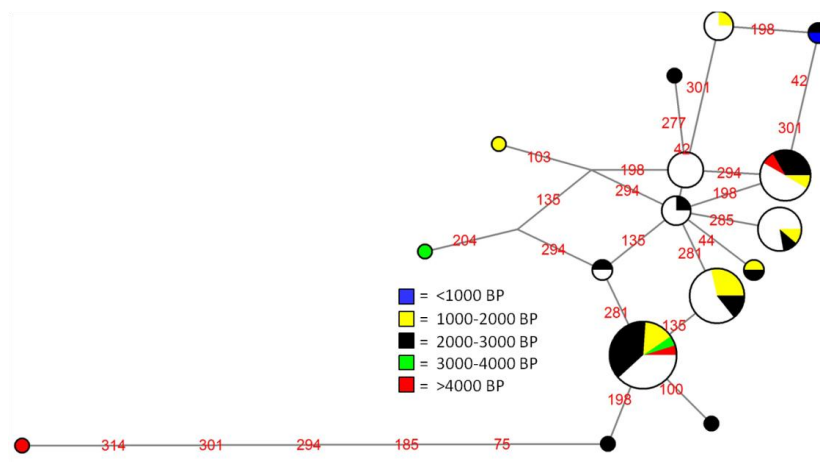
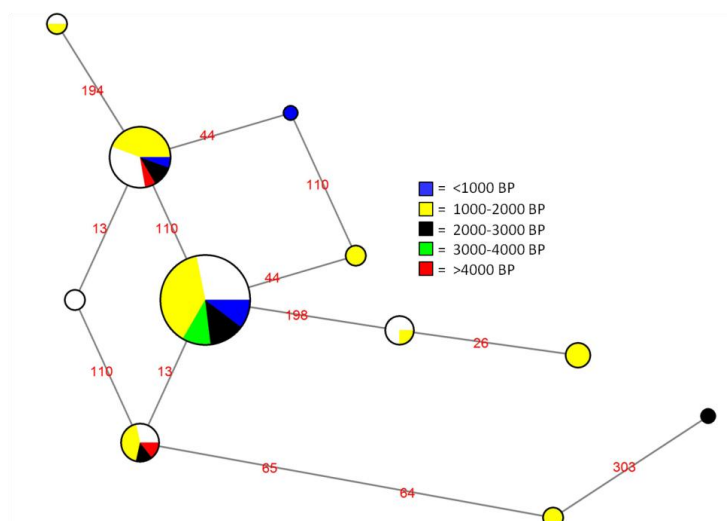


Figure 4.2e: *Emeus crassus*



4.2.3 Demography

To investigate the demographic history of the four moa populations, DNA sequences with associated radiocarbon dates were analysed in BEAST ver. 1.5.4 (Drummond, Rambaut, 2007). BEAST attempts to establish the most likely genealogy by use of a coalescent-based model in a Bayesian framework. User-defined prior values for parameters such as population size and mutation rate can be incorporated into the model and these are updated through the MCMC to maximise the overall tree-likelihood. Hence, the method can be used not only to determine the most likely gene tree, but also to estimate a range of parameters associated with the evolutionary history. If individuals have been sampled from a small population, the genealogy should display a disproportionately high concentration of coalescent events (i.e. short branches) because the common ancestor is likely to be found at only a few generations back. Conversely, individuals sampled from large populations show low concentration of coalescent events (longer branches) since the ancestral relationship among the sampled individuals is likely to be more distant. The expected waiting times (branch lengths between coalescent events) between two individuals is $2N_e$ generations, and this association is utilized in the generalised Bayesian Skyline Plot (BSP) to estimate effective population sizes (N_e) through time (Shapiro et al., 2004; Strimmer, Pybus, 2001).

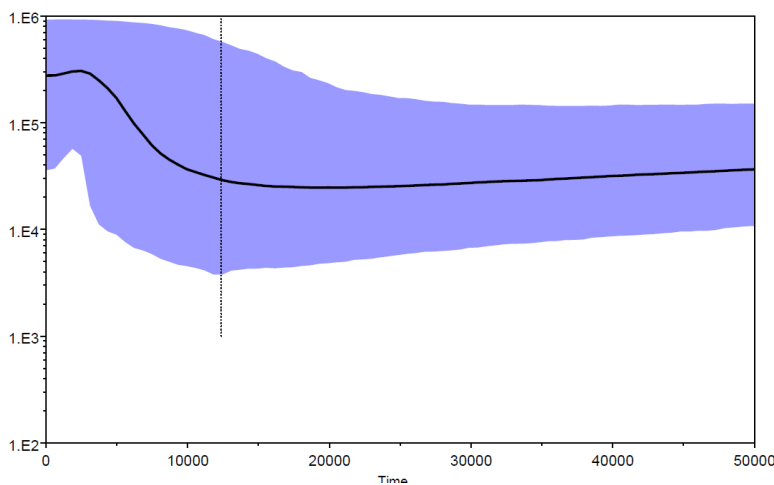
To maintain chronological accuracy of the coalescent events in the genealogy, only sequences with associated radiocarbon dates were included in this analysis. The four alignments were first imported into JMODELTEST ver. 0.1.1 (Posada, 2008) to establish the most appropriate substitution model for these data. By applying the Akaike Information Criterion (AIC) (Akaike, 1974), implemented in JMODELTEST, a HKY+I+G model was favoured for all four datasets. Prior values on parameters associated with this substitution model were incorporated as estimated in JMODELTEST using wide prior distributions. A relaxed molecular clock was applied with a prior mutation rate of 8.7×10^{-8} per site per year (as estimated for moa in Bunce et al. 2009). The mutation rate was incorporated as a mean value in a normal distribution (SD = 0.0005). The prior value for effective population size (N_e * generation time) was set to 10^3 with a uniform distribution of 100 to 10^6 . This can be considered a broad prior, but this was necessary to cover previous estimates of moa population sizes (Gemmell et al., 2004; Holdaway, Jacomb, 2000). Ten sampling groups were used to smooth the BSP. Initial test runs evaluated in TRACER ver. 1.5 (Rambaut, Drummond, 2007) indicated that long MCMC runs were necessary to reach chain convergence and effective sample sizes (ESS) above 200, as recommended by the authors (Drummond et al., 2007). Hence, for the full analyses, the MCMC was set to generate 10^8 trees, sampled every 10^4 states.

The resulting skyline plots for the four sympatric moa populations were visualised in TRACER, after removing the first 10 % of the trees as burn-in (Figures 4.3a-d). In general, the plots showed large 95% highest posterior distributions (HPDs), which compromises the value for interpretation.

However, the signal for *Dinornis robustus* suggested an increase in effective population size commencing 12-14,000 years ago and continuing throughout the Holocene, before reaching a stable population size c. 2000 years before extinction (Figure 4.3a). The smaller HPDs, and hence perhaps more reliable demographic signal for *Dinornis*, is likely a reflection of a larger, more informative dataset (many radiocarbon dates, more genetic variability) compared to the other taxa. Therefore, a four-step procedure was applied to evaluate the strength of the signal in *Dinornis*:

1) First, many independent BEAST runs, using a range of different prior settings, were conducted; they all yielded BSP signals very similar to that shown in Figure 4.3a. Incorporation of an age-dependent DNA damage model, appropriate for some aDNA data (Rambaut et al., 2009), introduction of 50-year error margins on each individual radiocarbon age, or employing a GTR+I+G substitution model did not affect the BSP signal. Switching to a strict molecular clock had a marginal effect, in that it narrowed the HPDs slightly and increased the (mean) time to the root of the tree (Digital Appendix 12). In total, these analyses confirmed that the recorded BSP signal for the *Dinornis* data was robust.

Figure 4.3a: *Dinornis robustus*



Figures 4.3a-d: BSPs representing the four North Canterbury moa populations. Y-axis shows a product of effective female population size and generation time ($N_{ef} \times t$) with blue areas defining the 95% HPDs. X-axis is years before the youngest radiocarbon age (year 0) for each species respectively. The demography is estimated back to the first split (last coalescent event) on the gene tree, with the depicted timeframe defined by the mean of the posterior distribution. Dashed vertical line represents the 95% lower HPD of the age of the oldest split.

Figure 4.3b: *Pachyornis elephantopus*

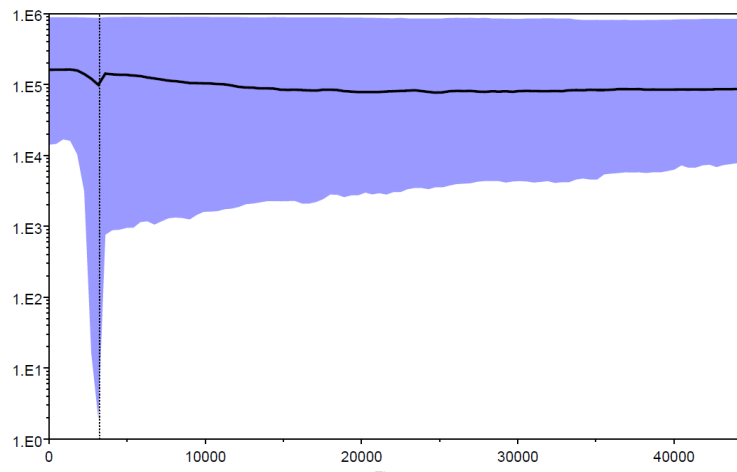


Figure 4.3c: *Euryapteryx curtus*

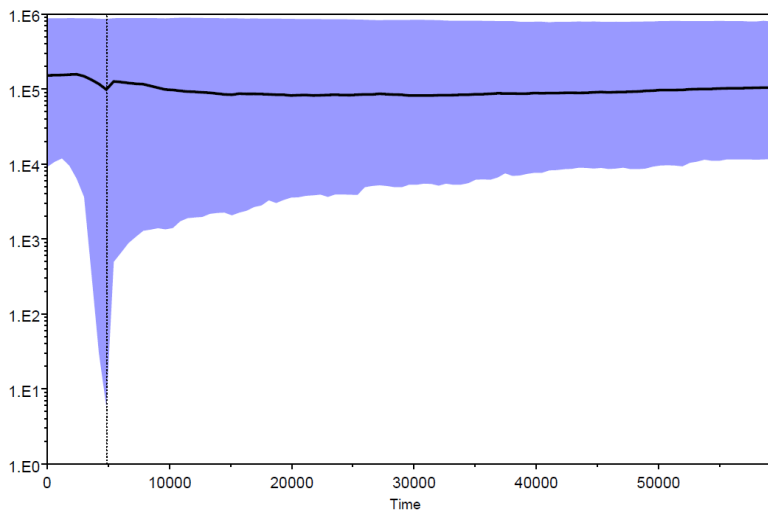
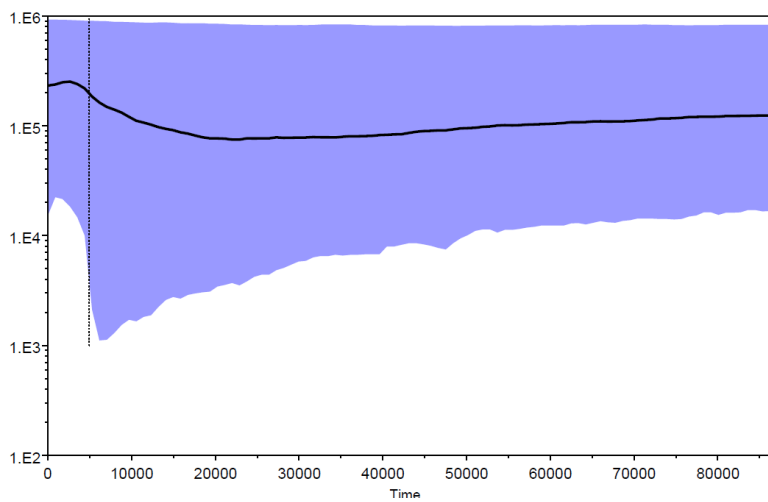


Figure 4.3d: *Emeus crassus*



2) Next, BEAST was run from an empty alignment. By applying the same user-defined model and prior settings, but excluding the DNA sequences, it was possible to test if the BSP signal was driven by the genetic information, or if it was simply an effect of prior settings (Drummond et al., 2007). The output from that run showed very small ESS values and a BSP signal that did not resemble the previous runs (Figure 4.4). This result confirmed that the original BSP signal observed for *Dinornis* (Figure 4.3a) was not simply a function of the user-defined settings.

3) In step three, two runs were performed as before but with the associated radiocarbon dates randomly shuffled amongst the DNA sequences, to assess if the BSP signal was driven by the genetic information rather than the ages. The latter situation can arise if the data lack sufficient phylogenetic information (Stiller et al., 2010). Although not identical to the results shown in Figure 4.3a, the BSPs with age-shuffling displayed comparable trends of a population increase (Figure 4.5). The HPDs were of the same magnitude, as was the mean posterior likelihood value ($\ln P \approx -724$), as in the BSP outputs from ‘unshuffled’ data. The similarities suggest that the unity of age and sequence data in *Dinornis* is only partly responsible for the BSP signal in Figure 4.3a. However, mean estimated substitution rates proved slightly higher in the two shuffled runs ($\mu \approx 3.3 \times 10^{-7}$ per site per year) than estimated from the original data ($\mu \approx 2.5 \times 10^{-7}$ per site per year); the posterior distributions largely overlapped, however, (Figure 4.6), which indicated that the differences may not be significant. A systematic approach involving more repetitions is necessary to statistically evaluate the differences between shuffled and unshuffled data, but the results presented here showed that the BSP signal in *Dinornis* should be interpreted with caution.

4) Lastly, a Bayes factor analysis was conducted to compare likelihood values derived from different demographic scenarios in BEAST. The BSP result for *Dinornis* was compared to three alternative and less complex coalescent models, assuming: (i) constant population size; (ii) exponential growth; and (iii) expansion growth⁷ respectively. For each model, the outputs from four independent MCMC chains (10^8 states sampled per 10^3 trees) were combined into a composite chain using LogCombiner ver. 1.5.4 (Drummond, Rambaut, 2007). TRACER was used to sample and compare marginal likelihood values. The results are shown in Table 4.3. The Bayes factors strongly supported a constant population size or an exponentially increasing population size in favour of the BSP and the expansion model. Bayes factors could not discriminate between a constant population size and an exponentially increasing population. From a parsimonious viewpoint, we cannot reject the simplest demographic scenario of a constant population size. Similarly, Bayes factors were calculated and compared for different demographic models within each of the three other moa species, but likely because of a lack of resolution power, the different demographic models provided almost identical likelihood values.

⁷ A model of increasing growth rate, relevant if, for example, the population is expanding into new favourable habitat

A few mtDNA, CR sequences with associated radiocarbon dates generated outside this study were available for each species, but the inclusion of those data did not change the BSP signals or the likelihood values presented here.

Figure 4.4: BSP from empty alignment

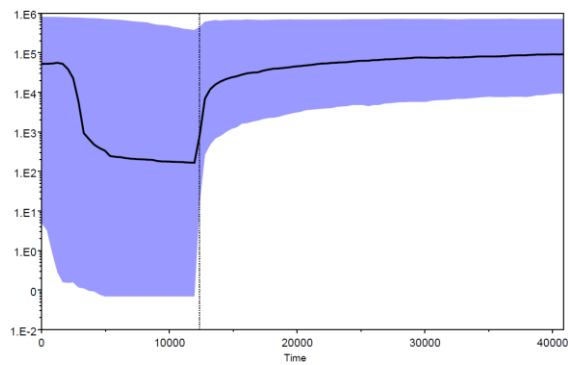


Figure 4.4: Bayesian Skyline Plot representing *Dinornis* but sampled from an empty alignment. The signal is different from the BSP in Figure 4.3a, showing that the recorded signal in Figure 4.3a is not simply a result of user-defined settings.

Figure 4.5: BSPs with radiocarbon-ages shuffled amongst DNA sequences

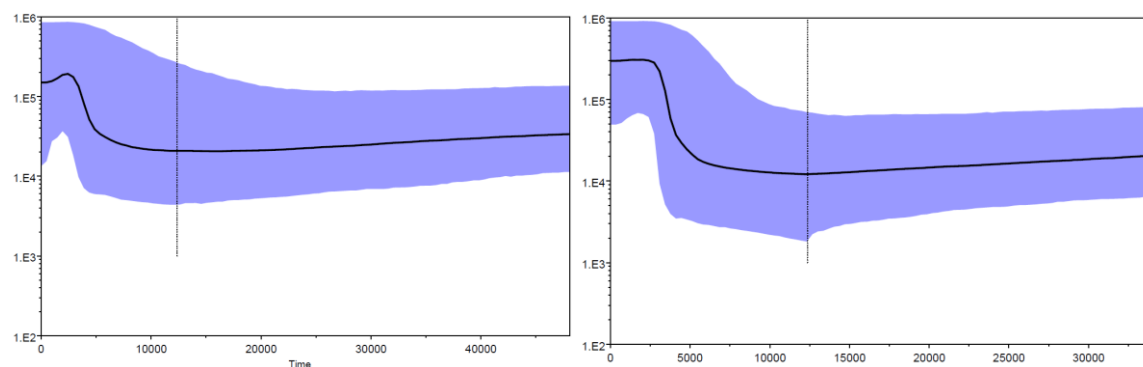


Figure 4.5: BSPs representing *Dinornis* but with shuffled radiocarbon-dates. The signal is comparable to the original BSP in Figure 4.3a, perhaps indicating that age-data is driving the signal rather than genetic information.

Figure 4.6: Substitution rates

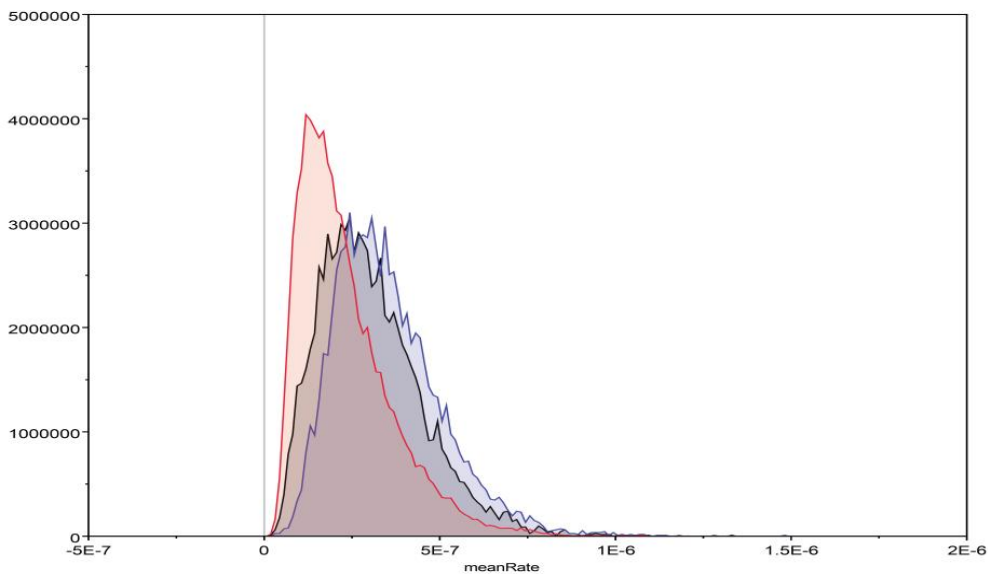


Figure 4.6: Posterior distributions of substitution rates as estimated in n BEAST based on *Dinornis* data. The red distribution represents the original data with a mean estimated μ of 2.5×10^{-7} per site per year, whereas the two other are based on shuffled radiocarbon ages, with means of 3.0×10^{-7} and 3.5×10^{-7} , respectively.

Table 4.3: Marginal likelihood values (including standard errors from 1000 bootstraps) and Bayes factor (BF) calculations conducted in TRACER based on *Dinornis* data. The BFs correspond to row-by-column comparisons of the demographic models.

	$\ln P(\text{model} \text{data})$	BSP	Constant	Exp. growth	Expan. growth
BSP	-744.0 ± 0.22	—	0.0	0.0	0.1
Constant	-737.8 ± 0.18	480.3***	—	0.8	27.8**
Exponential growth	-737.6 ± 0.18	644.0***	1.3	—	37.3**
Expansion growth	-741.2 ± 0.21	17.3*	0.04	0.0	—

*denotes weak, ** denotes strong, and *** denotes very strong evidence in favour of the row model according to Goodman (1999).

4.2.4 An estimate of population size

Despite the uncertainties inherent in the demographic analyses, they do provide some basis for estimates of moa population sizes – in particular when considering that the only previous genetic estimate was based on a much smaller dataset: Gemmell et al. (2004) aligned 36 *Dinornis robustus* control region sequences compiled from GenBank, and estimated an effective population size (N_e) of

the entire South Island population, using first a classical theoretical correlation between genetic diversity, mutation rate, and N_e , followed by a three-step conversion protocol (outlined in Roman, Palumbi, 2003) to retrieve an estimate of the population census size (N_c). A similar assessment can be based on N_e values derived from the BEAST analyses:

The Y-axis on the BSP represents a compound estimator ($N_{ef} \times t$) of effective female population size (N_{ef}) multiplied by generation time (t) (Drummond et al., 2007). From this compound estimator, it is possible to derive an estimate of the census population sizes N_c at different times. As the simplest demographic model of a constant population size could not be rejected, the posterior distribution from this model was used to estimate N_c for the South Island giant moa (*Dinornis robustus*). The mean $N_{ef} \times t$ from the log-combined output file of four runs was 215×10^3 with 95% HPDs from 15×10^3 to 733×10^3 . To estimate N_{ef} , a moa generation time (t) must be established. Analyses of ontogenetic growth lines in moa long bones suggest that moa exhibited a K -selected life history with delayed maturity (Turvey et al., 2005; Turvey, Holdaway, 2005), so t -values for extant ratites are probably invalid for moa. Lines of arrested growth (LAGs) were observed among emeids up to nine years of age (Turvey et al., 2005), which suggests that full sexual maturity may have been delayed to at least age 10. For *Dinornis*, the data suggest a shorter maturation (Turvey et al., 2005), which indicates a generation time shorter than 10 years. However, a matriarchal territorial system as suggested in Chapter 2 (Section 1) could mean that only large fully-grown females reproduced. On this basis, an average age at first breeding of 10 years for a giant moa female seems the best available estimate (R.N. Holdaway, pers. comm.).

By applying this conversion factor, the result of the BEAST analysis suggests an effective female population size (N_{ef}) for the North Canterbury *Dinornis* of 21,500, with 95% errors (based on HPDs): 1500-73300 individuals. These numbers have been rounded to the closest thousand in Table 4.4. Although derived by a different approach using a different and much larger dataset, these numbers are comparable to the estimated N_{ef} in Gemmell et al. (2004) (Table 4.4).

The three-step conversion protocol (N_{ef} to N_c) from Gemmell et al. (2004) is as follows:

Conversion 1) N_{ef} was converted to the total effective population size by including the fraction of reproducing males in the population. Gemmell et al. (2004) used the 1:1.4 male to female ratio estimated in Bunce et al. (2003). Because this ratio was derived from a sample of moa material from a range of fossil sites (including caves), it may not display the same gender bias as the sites dealt with here (see Chapter 2, Section 1). On the other hand, the 1:1.4 ratio was not based on accurate molecular sexing but on a morphological re-assessment of the bones, assuming that the smaller individuals were males. Reflecting the inaccuracy in using morphology as a gender identification, Bunce et al. (2003) estimated a 1:1 sex ratio in *Dinornis* from BHV, whereas molecular sexing

showed a 1:1.8 ratio in favour of *Dinornis* females at this site (Chapter 2, Section 1). Conversion 1 was here performed with both the 1:1.4 ratio applied by Gemmell et al. (2004) and the 1:2.2 overall sex-ratio estimated from BHV (31% males, Table 4.4), which was considered more likely to reflect that in a natural moa population (Chapter 2, Section 1).

Conversion 2) In the next step, N_e was converted to the census size of the moa population. Gemmell et al. (2004) multiplied N_e by a factor of 2, using the same argument as Roman and Palumbi (2003) that N_c is often approximately double N_e under constant population sizes (Nunney, 1991). The conversion factor can be considered conservative given that genetic assessments of N_c/N_e have often yielded values closer to 10 in natural populations (Frankham, 1995; Mace, Lande, 1991; Nunney, 2000; Nunney, Campbell, 1993). For comparison, and because we lack information on population dynamics and mating structure in moa, the same conversion factor ($N_c/N_e = 2$) was applied here (Table 4.4).

Conversion 3) The final step accounts for the fraction of non-reproducing juveniles in the population, which must be added to get the total standing population N_c . Gemmell et al. (2004) used an estimate of 16 % juveniles, as previously estimated from the Pyramid Valley fossil record by Worthy and Holdaway (2002). However, as shown in Chapter 2 (Section 1), this site has preserved a highly biased sample of the surrounding moa community, with an extreme excess of adult females. Conversion 3 was therefore carried out using both a value 16 % juveniles, but also using the value 41% apparent at Bell Hill Vineyard (Table 4.4), which may better reflect the situation in the living population (Turvey, Holdaway, 2005; Chapter 2, Section 1).

Table 4.4: Estimates of population sizes of *Dinornis robustus* during the late Holocene. Gemmell et al. (2004) estimated the effective female population size (N_{ef}) based on 36 mtDNA control region DNA sequences and two different mutation rates (μ) from the literature. The BEAST estimates were based on $N_{ef} \times t$ results from the BEAST-runs, assuming a constant population size and a generation time (t) of 10 years. BEAST estimate1 applies the same conversion factors as Gemmell et al. (2004) to calculate the census size (N_c), whereas a different ‘updated’ set of conversion factors, discussed above, were applied in BEAST estimate2.

	$N_{ef} \times 10^3$ (95% CI)	conversion info			$N_c \times 10^3$ (95% CI)
		% males	N_c/N_e	% juv.	
Gemmell et al., low μ	117 (79-170)	42	2	16	479 (321-694)
Gemmell et al., high μ	25 (17-36)	42	2	16	102 (68-147)
BEAST estimate 1	22 (2-73)	42	2	16	90 (8-299)
BEAST estimate 2	22 (2-73)	31	2	41	108 (10-358)

The population size estimates derived from the BEAST runs suggest that the 87 sequences included in this study were drawn from a *Dinornis robustus* population of about 10^5 individuals and considerably smaller than the estimate from Gemmell et al. (2004) (Table 4.4). The error margins are large, however, (despite the comparative large sample in aDNA terms), likely because of limited genetic information available.

4.2.5 The bigger picture – phylogeography

The North Canterbury moa data must also be examined in a larger genetic context. All relevant moa mtDNA CR-sequences were downloaded from GenBank and aligned with the North Canterbury data presented here. Only a few GenBank sequences had associated radiocarbon dates, so the phylogeographic analyses were done without using any age information (tip dates). It was performed solely to identify the presence of any genetic structuring in the data, whether temporal, spatial, or something else. Only sequences including the 340 bp region, without gaps, and ambiguous sites, and having a recorded provenance, were included in the analyses. These criteria yielded 34 *Dinornis robustus*, 14 *Pachyornis elephantopus*, 41 *Euryapteryx curtus*, and 6 *Emeus crassus* sequences, which were aligned with the North Canterbury moa sequences and trimmed to the shortest sequence in GENEIOUS. The four alignments (including outgroups) are given in Digital Appendix 1. The phylogenetic relationships within each species were analysed in BEAST based on a Yule model of speciation, and a strict molecular clock. The Akaike criterion in MODELTEST favoured the HKY + I + G model and the MCMC was set to run for 10^8 iterations, sampling every 10^4 states. As before, the signals were examined for convergence and sufficient ESS values in TRACER. TreeAnnotator ver. 1.5.4 (Drummond, Rambaut, 2007) was used to identify the maximum a posteriori (MAP) trees, with branch lengths defined by the mean of each node distribution, and the trees were visualised in FigTree ver. 1.2.2. (<http://tree.bio.ed.ac.uk/software/figtree>).

The MAP trees are shown below in their entirety, with all branches, tips and node values displayed. Although this level of detail might seem unnecessary, important information is contained in the sequence names (locality and age), and collapsing the branches would exclude this information. Rather than collapsing branches, colour-coding was applied to assist the reader in the following outlined interpretations of the observed genetic structuring (Figures 4.7a-d).

Dinornis robustus (Figure 4.7a)

In total, 121 *Dinornis* sequences were aligned and analysed, including the 87 (71.9 %) generated in this study. The MAP tree showed several patterns. Although the relative positions were not clearly defined, there were three deep-rooted and highly-supported monophyletic clades in the tree: two individuals from Central Otago (A, node value 1) and two individuals from Honeycomb Hill Cave in the northwestern South Island (B, node value 1) constituted the two basal clades respectively, although the MAP tree was not informative on which split happened first (depicted node value: 0.26). The third clade (C, node value 0.99) included all the North Canterbury individuals, a group of individuals from South Canterbury and Northern Otago (mainly Kapua and Enfield, North and South of the mouth of the Waitaki River, respectively), and then sequences from the northwestern South Island (red). Although founded within clade C, most individuals within the “red region” formed a distinct clade (D, node value 0.94). The exceptions were three individuals from the red area (two from Takaka Hill, one from Maximus Cave), which fell outside this clade, and instead amongst the “green region” of the tree. Within this green Canterbury/Otago clade, some structure was apparent (although not geographically defined), in that a large group of North Canterbury individuals was relatively well supported (E, node value 0.87): this is the same clade that constituted haplogroup A in Figure 4.2a.

Pachyornis elephantopus (Figure 4.7b)

In total, 45 *Pachyornis* sequences were aligned and analysed, including the 31 (73.8%) generated in this study. Three well-supported clades were apparent: an East Coast clade (A, node value 1) from North Canterbury to Pareau in East Otago, a Southland/Stewart Island/Central Otago clade (B, node value 1) and lastly; an East Otago clade (C, node value 0.98) embedded within clade A. Interestingly, two of the lineages (B and C) have been sampled from Chatto Creek, which is geographically located in between.

Euryapteryx curtus (Figure 4.7c)

Of 123 *Euryapteryx* sequences aligned and analysed, 82 (66.7%) were generated in this study. Individuals from both islands were included, because it has been shown previously that there was no North-South Island monophyly in this genus (Bunce et al., 2009). In contrast to *Dinornis*, no deep-rooted clades were observed among *Euryapteryx* and the relative order of branching events was unclear. However, some geographic structure was evident nonetheless. Although not strong, there was some support for a North Canterbury (green) clade (A, node value: 0.40), except for one distinct Rosslea individual which was outside this clade. This was identified as the single, very distinct,

haplotype observed in the North Canterbury network above (Figure 4.2d). The sub-structure along the East Coast, observed in *Pachyornis*, was mirrored here in *Euryapteryx*, with the birds from East Otago (blue) representing a very distinct genetic unit (B, node value: 0.99), including a single individual from Central Otago (Glendhu Bay). Individuals from the northwestern South Island (red) separated as a distinct clade (C, node value: 0.99) as was also apparent in the *Dinornis* tree. The genetic affiliation of a single individual from Southland (yellow) was inconclusive, possibly because it constituted its own “group”; several unique mutations were represented in this individual. Interestingly, all North Island individuals appeared together as a monophyletic group in this MAP tree, but there was no support for this node (0.05), which confirmed that some of the geographically-defined South Island clades (B and C) was much better supported than the North Island grouping as a whole. Clearly, there was genetic structuring *within* *E. curtus* population on the North Island, but the overall relationships between the clades in this taxon remain cryptic.

Emeus crassus (Figure 4.7d)

In total, 87 *Emeus* sequences were aligned and analysed, including 81 (93.1%) generated in this study. Because of the low genetic variability in this species, and because of the small number of available GenBank sequences, little genetic structure is apparent. However, visual examination of the alignment, showed that North Canterbury individuals were not genetically differentiated to South Canterbury individuals (Kapua) and to the single included *Emeus* from Southland (Castle Rock), hence only one colour appear in the tree. Despite limited data, it seems that the Eastern moa gene pool was more homogenous than the other moa.

Figure 4.7a: *Dinornis robustus*

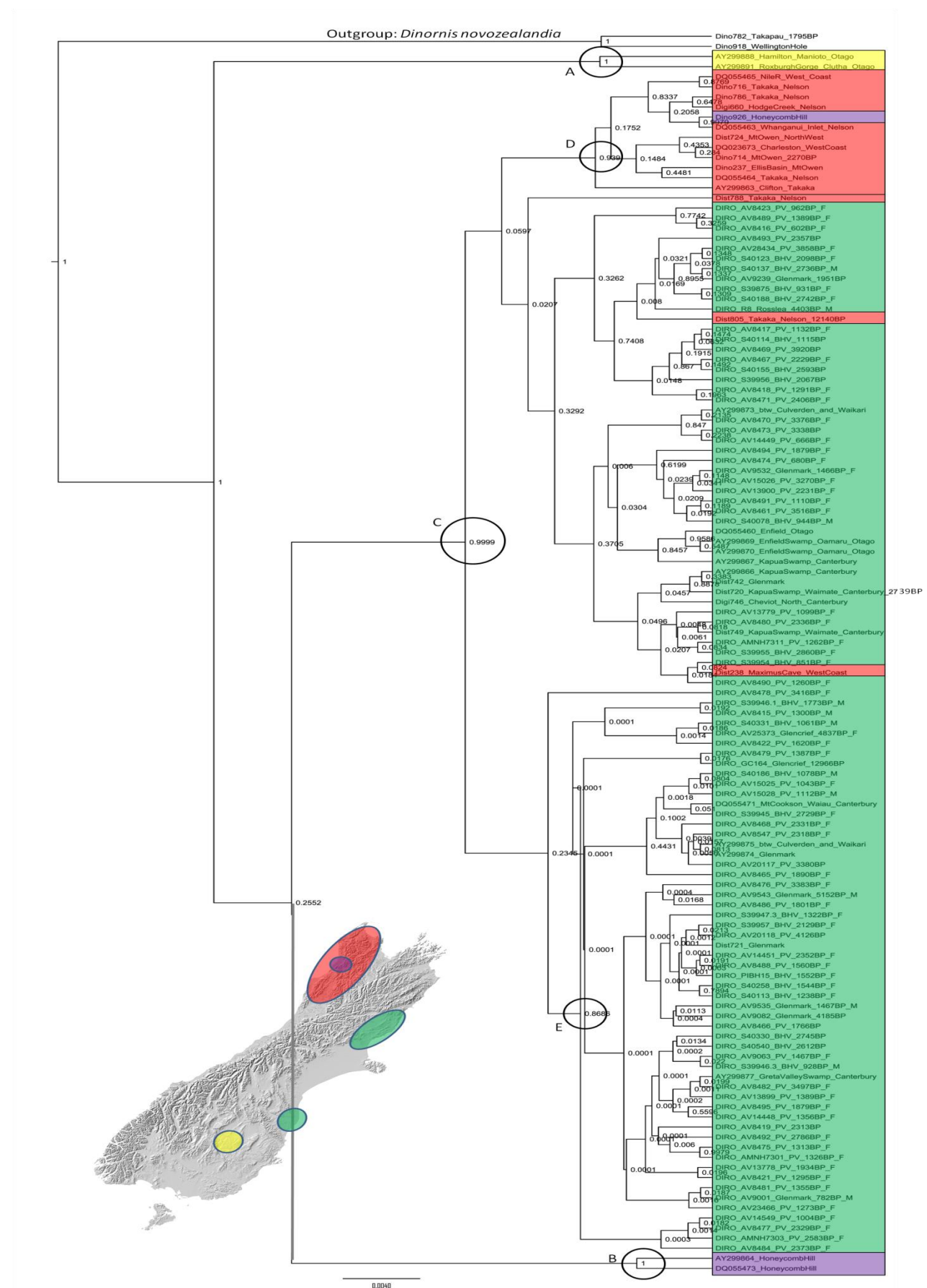


Figure 4.7b: *Pachyornis elephantopus*

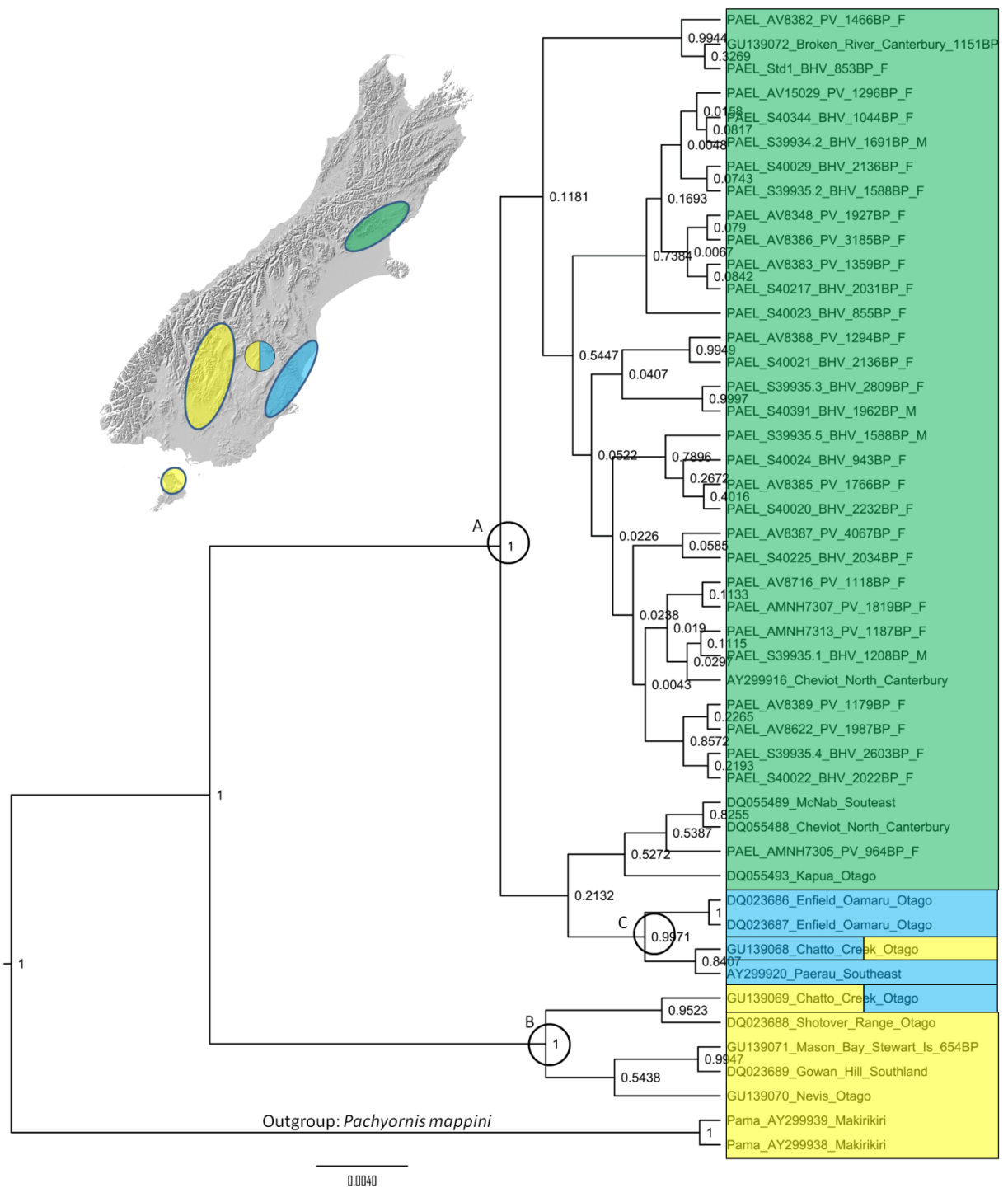


Figure 4.7c: *Euryapteryx curtus*

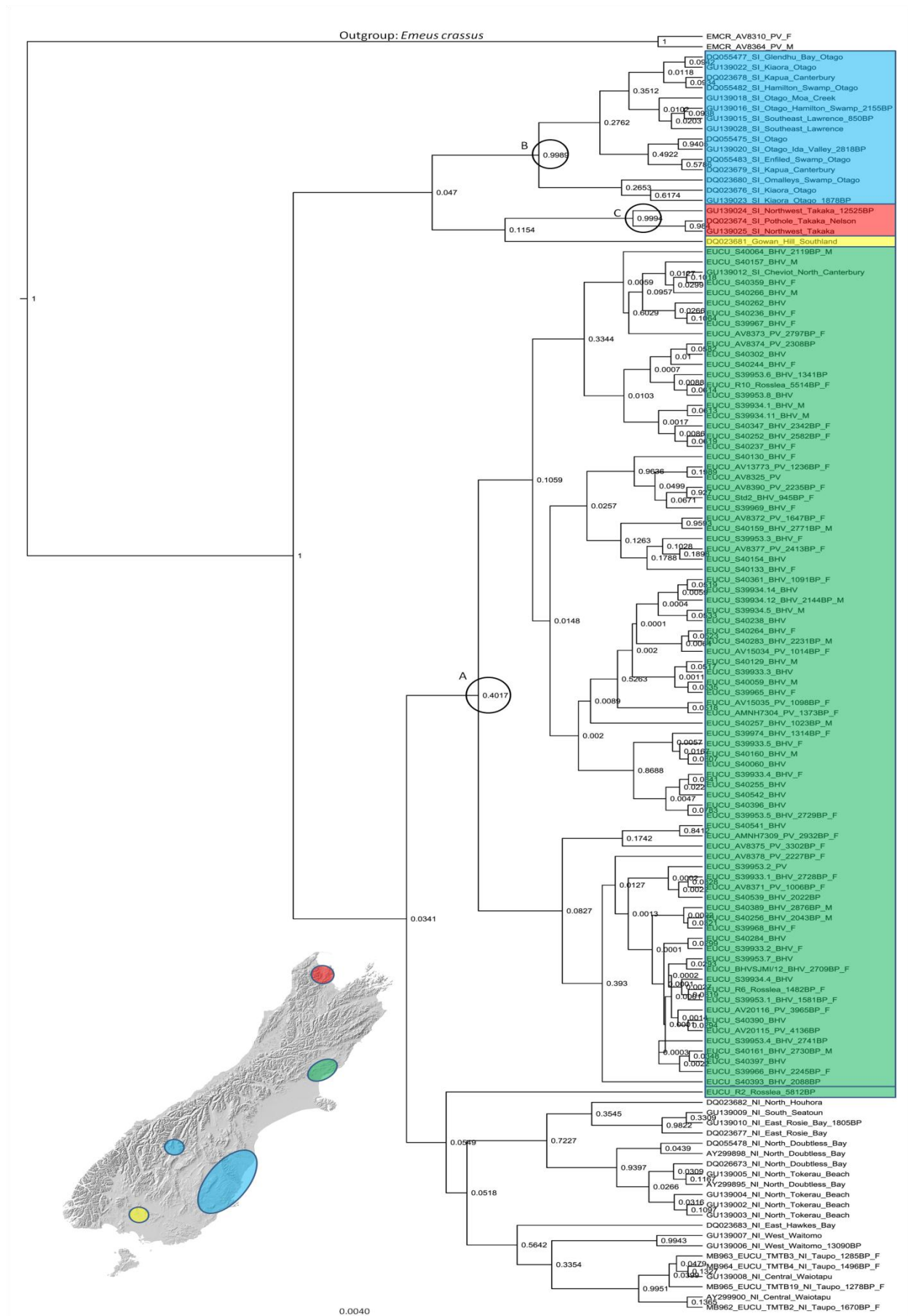


Figure 4.7d: *Emeus crassus*

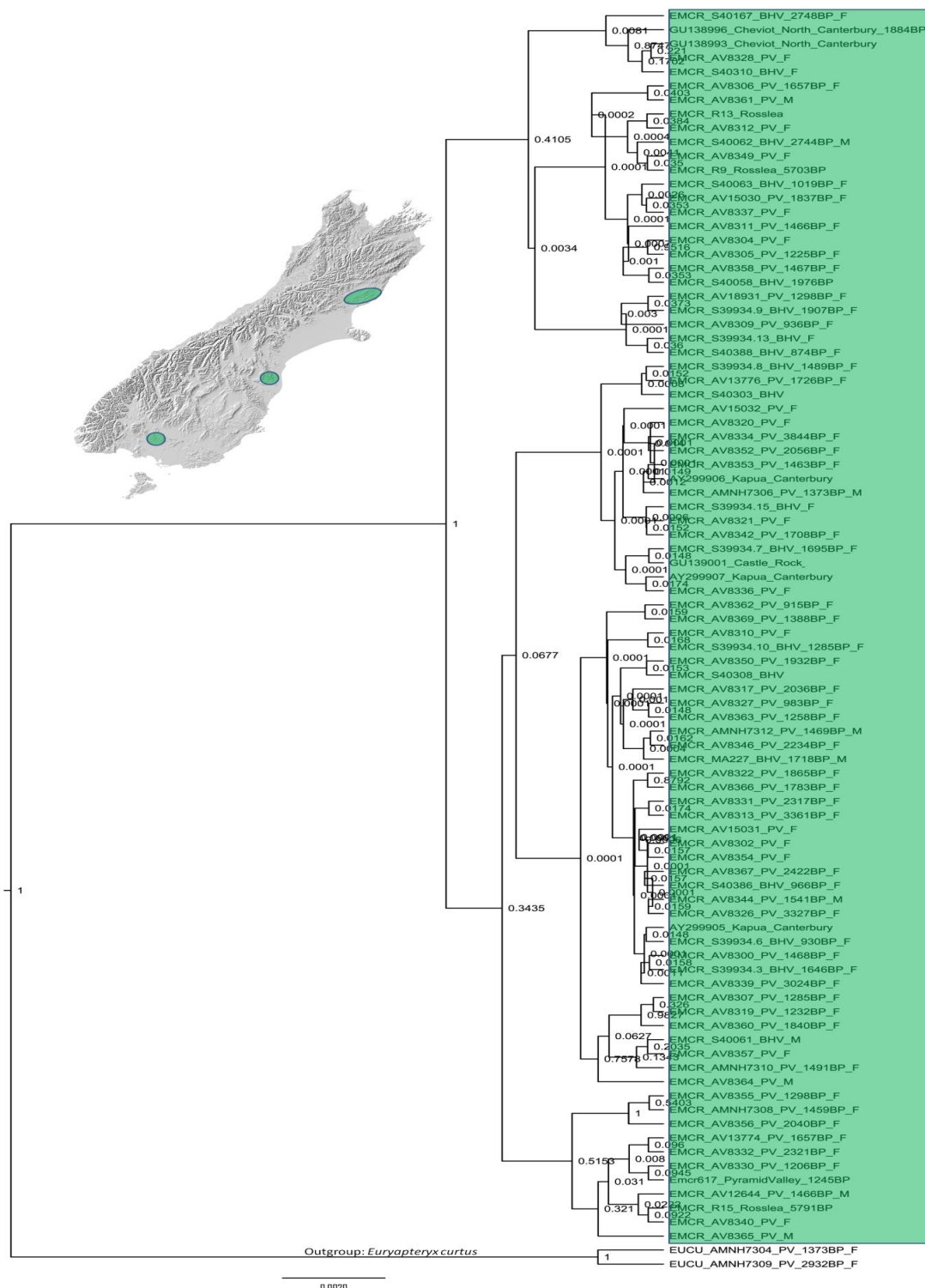


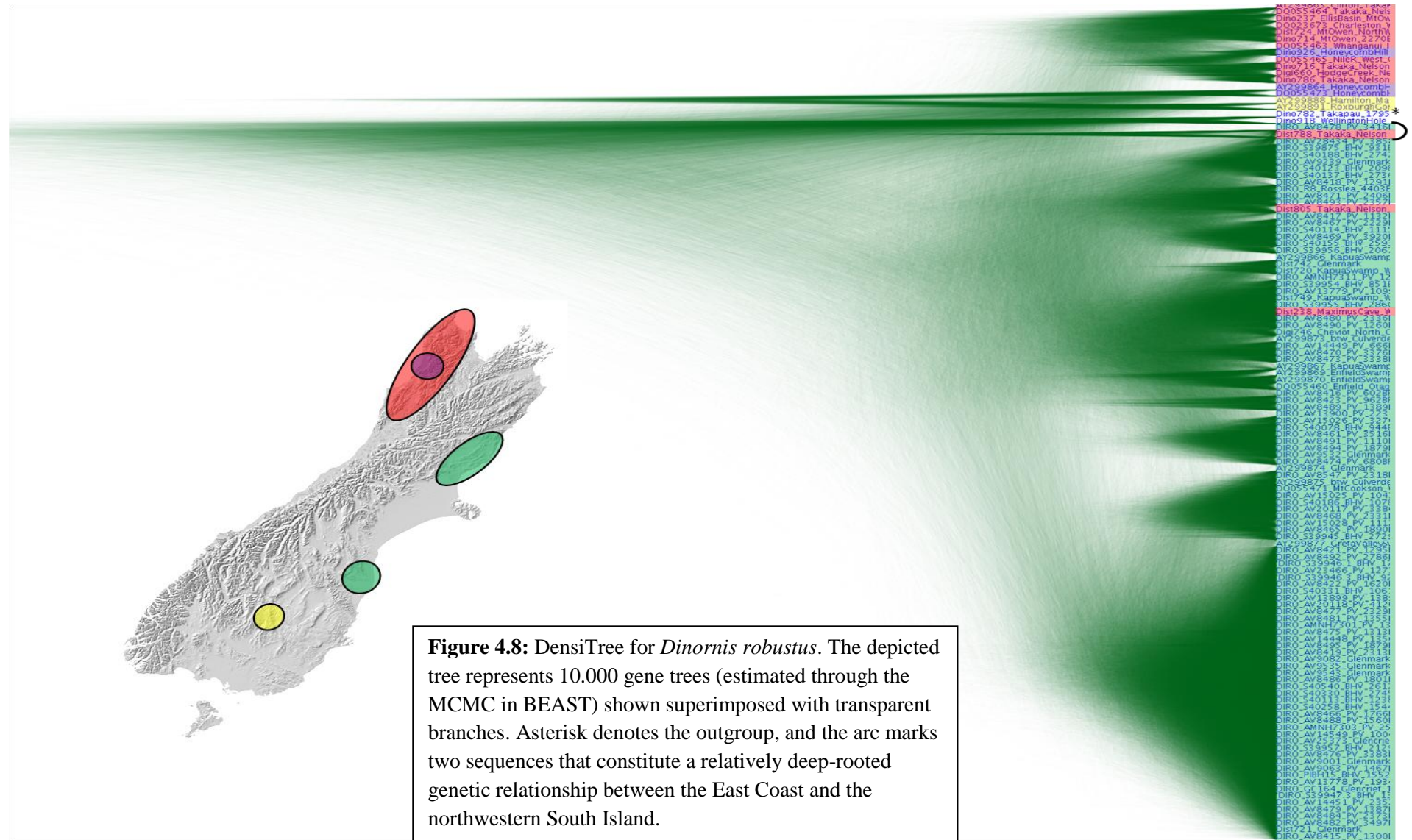
Figure 4.7a-d: MAP trees for the four moa species based on ~340 bp mtDNA (control region). Colour coding of each sequence name has been applied to assist in the interpretations as outlined in the text. Node values display proportion of trees sharing the topology (out of 10000 trees generated). No age information was used in the analyses so branch lengths represent raw genetic distances (substitutions per site). DIRO = *Dinornis robustus*, PAEL = *Pachyornis elephantopus*, EUCU = *Euryapteryx curtus*, EMCR = *Emeus crassus*, PV = Pyramid Valley, BHV = Bell Hill Vineyard, M = male, F = female. Associated radiocarbon ages are shown in years BP. In *Euryapteryx* tree: NI = North Island.

4.2.6 Assessing the robustness of the topology - a DensiTree analysis of *Dinornis robustus*

Finally, the *Dinornis* output file from BEAST was examined with DensiTree ver. 1.4, which represents a novel, qualitative approach for assessing the robustness of the phylogenetic information in data (Bouckaert, 2010). BEAST uses MCMC to many trees within a certain (partly user-constrained) distribution of all possible states in tree-space. However, it is conventional to display just a single MAP-tree as above, which implies that the topological information present in the remaining sampled trees (here 9999 trees) is ignored. DensiTree provides a method for visualising large tree sets by drawing them all, superimposed, with transparent lines. Areas where many trees agree in their topology and branch length will be densely coloured, whereas areas of little consistency across trees are seen as a smear of faint lines.

Overall, the DensiTree for *Dinornis* (Figure 4.8) resembled a poorly resolved topology. Branch lengths between different varied significantly and, as already seen from the MAP-tree, the branching order was not easily resolved either. Hence, large areas of the DensiTree tree were seen as a faint network of different node heights and basal branches projecting in different directions. However, although the branching order was unclear, the tree confirmed a considerable level of genetic structure in this species, in terms of both the wider geographic structure, as discussed above, but also including distinct mitochondrial lineages within the North Canterbury population. The large, dense, region in the lower part of the tree was, for example, equivalent to haplogroup A on Figure 4.2a. The deep splits consisted of (apart from the outgroup) two Honeycomb Hill individuals (purple) and two from Central Otago (yellow). An interesting feature revealed in the DensiTree but not visible in the MAP-tree, was the relatively well-defined genetic relationship between a particular North Canterbury *Dinornis* (“C”, on Figure 4.2a) and another single individual from northwestern South Island (Dist788, Takaka). The relationship may represent a genetic link between *Dinornis* in the two geographic regions.

Figure 4.8: DensiTree – *Dinornis robustus*



4.3 Discussion

4.3.1 Age

As discussed in Chapter 1 (Section 3) and in Chapter 2 (Section 1), a few radiocarbon ages were available from the five North Canterbury sites before this project. The available ages indicated that the two main sites, Pyramid Valley and Bell Hill Vineyard were late Holocene fossil deposits, providing an unparalleled, comparative setup for studying moa populations. This has now been confirmed by one of the largest ^{14}C -chronologies on extinct taxa that have ever been assembled. Samples from three other sites in immediate vicinity, Rosslea, Glenmark and Glencrieff were incorporated at a later stage, with the aim of providing greater temporal depth to the demographic analyses.

From a molecular viewpoint, the radiocarbon dates serve as individual coordinates in time, providing a framework for analysing and interpreting the genetic variation. The radiocarbon age confirmed the relatively close temporal proximity between fossils from different taxa, and different sites, respectively. This is most important, as it has allowed the genetic results to be interpreted in the context of moa populations rather than as small, isolated, “snapshots” of gene pools widely separated in time and space.

There is a current consensus that Polynesians settled in New Zealand late in the 13th century, with Wairau Bar, at the northern end of the South Island, being the oldest securely-dated archaeological site (Higham et al., 1999; Wilmschurst et al., 2008). The here presented ^{14}C chronology suggests that moa deposition in natural North Canterbury sites ceased c.1350 CE (i.e. 1950 CE - 600)⁸. Several archaeological sites in the region show that moa were hunted extensively in Canterbury (summarised in Anderson et al. 1989). At Monck’s Cave, (at the junction of Banks Peninsula and the coast to the North), there are no secure signs of moa consumption, suggesting that they were no longer available when the cave was occupied. Holdaway & Jacomb (2000) dated the archaeological layers in and around Monck’s Cave and concluded that humans occupied the site from the mid- to late-14th century CE in an environment without moa. This conclusion has now been supported by an extensive chronology based on fossils from natural North Canterbury sites, showing that the natural deposition of moa ceased in the region around 1350 CE, which is therefore likely to be an accurate estimate of local extinction time.

Only four moa examined here, were dated to between 800 and 600 BP and hence *potentially* represent the moa populations encountered by the early Polynesian settlers. Therefore, even if the rapid

⁸The convention for reporting radiocarbon dates is to reference according to 1950 CE (Common Era = AD), which was the last year before atmosphere nuclear testing changed the reservoir levels of ^{14}C in the atmosphere. Therefore, radiocarbon dates are shown as BP (before present), with “present” being 1950 CE.

extinction process left any genetic traces, the material available here would be unlikely to have recorded it. The youngest calibrated ^{14}C age in the North Canterbury corpus of dates had a median calibrated age of 602 years BP (95% CI = 545-662 BP). This *Dinornis robustus* from Pyramid Valley, is now the youngest moa ever dated from a natural site, and is about 50 years more recent than the *Dinornis* reported by Worthy (2002).

A potentially important feature of the PV and BHV date series, is the apparent *pulsed* deposition, suggesting that incorporation efficiency alternated between Pyramid Valley and Bell Hill Vineyard. Different factors, perhaps related to climatic fluctuations may be responsible for these patterns and the data will be explored in detail elsewhere.

4.3.2 Genetic diversity

It was suggested in the earlier molecular studies that moa displayed a low level of mtDNA diversity (Cooper, Cooper, 1995; Cooper et al., 1992), and this trend is perhaps repeated in other New Zealand birds such as kiwis (*Apteryx* spp.) and wrens (Acanthisittidae) (Baker et al., 1995; Cooper, Cooper, 1995; Shepherd, Lambert, 2008). A comparison of results in this study with genetic studies of other ratites did not clarify this matter (see Chapter 5). Regardless, Cooper and Cooper (1995) hypothesised that the Oligocene drowning (see Chapter 1, Section 1) was responsible for loss of genetic diversity in the biota, because of severe demographic bottlenecks. However, given that moa and kiwi may have arrived later to New Zealand (e.g., Phillips et al., 2010) with a much more recent radiation (Bunce et al., 2009), it seems possible that more recent events, such as repeated glacial cycles throughout the Pleistocene, and frequent catastrophic volcanism in the North Island, were responsible for the observed intraspecific genetic patterns.

In particular, *Emeus crassus* displayed very little variability, with most sequences being identical in the amplified CR fragment. Only 11 haplotypes were identified among 81 sampled individuals (Table 4.2, Figure 4.2e). This is an indication that this species in particular has undergone a genetic bottleneck as set out in the introduction to this chapter. This could, for example be a consequence of prolonged isolation of a remnant population in a glacial refuge during the Otiran glaciation, followed by a re-colonisation of the Eastern lowlands, when suitable habitat became re-established in the Pleistocene-Holocene transition (Bunce et al., 2009). The sheer abundance of fossils of this species (e.g., Holdaway, Worthy, 1997; Worthy, Holdaway, 1996; Worthy, Holdaway, 2002) suggests that there was a large population in the Eastern South Island despite the low observed genetic diversity. The topic is revisited in Chapter 5.

Two distinct haplogroups were present in the North Canterbury *Dinornis robustus* population, but there was also a single individual that seems to have belonged to a separate haplogroup, perhaps with a genetic relationship to the population in the northwestern South Island (Figures 4.2a and 4.8). However, the haplotype networks showed that genetic structures within each species were not defined by time or space. No obvious clustering could be identified, which implies that the three lineages in *Dinornis* reflected old maternal lineages within the same population.

The signal could be interpreted as the North Canterbury *Dinornis* carrying genetic remnants from two separate sub-populations. The general lack of structure with relevance to fossil site confirmed an expectation that the North Canterbury deposits were not sampling different moa populations, at least not to an extent that could be revealed by the control region data. Furthermore, no sudden shifts in haplotype frequencies through time were revealed in any of the four taxa. The same haplotypes were present throughout the geologic time covered by the data, perhaps suggesting that no major demographic disturbances influenced the four mitochondrial gene pools in the 4000 years leading up to the sudden extinction.

Microsatellite data allow more powerful analyses of intraspecific genetic structure, and these will be dealt with in Chapter 5.

4.3.4 Demography and population size

It is well established that a population expansion can lead to star-like gene-trees with an excess of rare haplotypes. Such patterns result because many lineages coalesce before the expansion when the population was small, and limited genetic drift towards the branch tips allow for the presence of rare haplotypes (Rogers, Harpending, 1992; Slatkin, Hudson, 1991). Star-like patterns have often been observed empirically, and interpreted as population expansions (Bowie et al., 2004; Dickerson et al., 2010; Magoulas et al., 2006; Mirol et al., 2008). Such a pattern is evident in the haplotype network for *Dinornis robustus* (Figures 4.2a and b). One interpretation of the *Dinornis* network could be that the North Canterbury region was a secondary contact zone in the Holocene and that lineage A represents a population that had expanded significantly before the secondary contact. The scenario is difficult to test, but simulation-based coalescent methods (Anderson et al., 2005; Cornuet et al., 2008) can model such complex scenarios and perhaps provide the necessary analytical framework.

Star-like patterns can emerge as results of processes other than population expansions. It has been shown that even in the absence of a demographic signal, heterochronous data can lead to significant over-estimates of mutation rates and genetic diversity, and also generate star-like tree structures with an excess of rare mutations (Debruyne, Poinar, 2009; Depaulis et al., 2009; Navascues et al., 2010). In

a coalescent framework, the increase in polymorphism observed in heterochronous data can be explained by longer coalescent times in non-contemporary lineages, providing a longer time period for mutations to occur (i.e. waiting time until lineages are contemporary + waiting time until the coalescent event) (Depaulis et al., 2009). Hence, if heterochrony is not accounted for, classical descriptive statistics applied for ‘normal’ homochronous data, will over-estimate genetic diversity. The implication is that the star-like pattern observed for *Dinornis* could potentially reflect mutations separating lineages through time rather than a high concentration of coalescent events happening at one point in time. Indeed, the star observed in Figures 4.2a contains all age groups and the associated rare haplotypes cover thousands of years, which suggests that the signal could be biased by this effect. However, arguing against the pattern being an artefact of heterochrony, is the absence of a star-like pattern in *Dinornis* lineage B, and in any of the other moa species. This is a good example of a benefit in applying a multi-species approach.

The Bayesian Skyline Plots incorporate the age information to increase the accuracy of the estimated genealogy through time: such analyses are strengthened rather than compromised by heterochronous data. Although the data suggested a population increase in *Dinornis* that began during Pleistocene-Holocene transition, the signal was not strong and it was not possible to reject a model of a constant population size through the investigated time period. More limited analyses were possible for the other species because of genetically less-informative datasets. The manual for BEAST states that “...the BSP should only be used if the data are strongly informative about population history, or when the demographic history is not the primary object of interest and a flexible coalescent tree prior with minimal assumptions is desirable” (Drummond et al., 2007). It is likely that the genetic diversity in the four datasets is too limited for generating accurate Skyline Plots. Moreover, it is not clear how powerful the BSP method is in situations of genetic stasis, nor whether the Bayes factor analysis suggested a constant population size because the data was insufficient or because the populations were indeed constant. Straight-line Skyline Plots with very narrow HPDs may be hard to generate and more simulation studies are needed to clarify these issues. The BSPs and the Bayes factor assessment indicated that the *Dinornis* population may have been increasing exponentially in size during the Holocene, and the haplotype network pattern for *Dinornis* suggested an increase in population size as well. The timing of this event fitted well with the ecological changes happening in the Pleistocene-Holocene transition.

A conservative interpretation of the BSP plots for all four species would be that there is no genetic evidence that moa went through a decline in the Holocene before their sudden extinction, as hypothesised by Gemmell et al. (2004). The interpretation of viable populations throughout the mid-to late Holocene is supported by observing continuing deposition at PV and BHV during that period (Figure 4.1).

The estimate of population size for *Dinornis robustus*, was problematic. Most conspicuous were the very large HPDs from the BEAST analysis, which were transferred through the various conversions to the final N_e estimate. The method employed by Gemmell et al. (2004) appeared to generate narrower confidence limits (Table 4.4), but the method is unlikely to provide reliable estimates of moa population sizes.

It is difficult, however, to adequately compare the two different approaches. The 36 DNA sequences used in Gemmell et al. (2004) represented moa fossils from many sites and widely separated geographic regions on the South Island. Their estimated population size therefore represents a hypothetical *Dinornis* population displaying panmixia through space and time. The BEAST estimates, on the other hand, are based only on North Canterbury genetic information (87 sequences), while accounting for genetic variability through time by incorporating radiocarbon dates. Hence BEAST provides an estimate of size of the North Canterbury *Dinornis* population, including all populations/sub-populations to which it was genetically connected throughout the Holocene. However, if for arguments sake, *Dinornis* panmixia is accepted, then the sampled North Canterbury gene pool should reflect the entire panmictic N_e – and therefore be directly comparable to the N_e estimate by Gemmell et al. (2004). This is a simple consequence of the assumptions underlying the latter study. By applying this hypothetical scenario, the BEAST-derived results, with a mean N_e around 10^5 (Table 4.4), suggest that the previous value of 4.79×10^5 individuals considerable over-estimated the size of the *Dinornis robustus* population.

Disregarding more trivial problems in applying a suitable mutation rate (μ) and appropriate N_e to N_e conversion factors, the method employed by Gemmell et al. (2004) is based on a simple correlation between genetic diversity and population size ($\theta = 2\mu \times N_{ef}$) which is highly questionable in this instance. θ is the population-scaled mutation rate, which can be interpreted as the average number mutations separating a sample of two sequences. The assumption of $\theta = 2\mu \times N_{ef}$ relies on selective neutrality and a Wright-Fisher model of random mating (e.g., Hedrick, 2000; Kimura, Crow, 1964) and does not account for population subdivision, or, for that matter, inflated estimates of genetic diversity resulting from heterochrony. Population subdivision result in genetic differentiation due to drift, and this will lead to an over-estimation of θ if all sequences are treated as if drawn from a single united gene pool. From a coalescent perspective, this means that average coalescent times become artificially prolonged, which is indicative of a large population according to the neutral theory. The species trees showed that there were clear geographic population subdivisions in *Dinornis*, so θ values estimated from a geographically combined dataset is likely to be overestimated, which also leads to an overestimate of N_{ef} .

The BEAST-derived estimates dealt with heterochrony by incorporating the radiocarbon dates; the assumption of panmixia was not breached, because the data had been derived from a very small

geographic area, and the data set was c. 2.5 times as large as that analysed by Gemmell et al. (2004) study. Even then, the confidence in the results was limited. Ironically, the lack of signal can probably be ascribed partly to sampling a population with lower genetic diversity, rather than to having data from large areas and wide time frames. Therefore, although the narrow temporal and spatial framework of the North Canterbury data represents its strength and uniqueness, it also imposes analytical limitations. Also the conversion factors represent sources of error, as each can significantly affect the N_e estimate. For example, if the generation time was five years rather than ten, all the BEAST-derived estimates of population sizes in Table 4.4 must be doubled.

In conclusion, the estimates derived from BEAST suggested a standing population of about 100,000 giant moa, but the error bars were wide, and moreover, the estimate is unlikely to represent the entire South Island, given the level of geographic subdivision. Although it is significantly smaller than the Gemmell et al. (2004) estimate of 479,000 South Island *Dinornis* individuals, the estimate is still larger than the 159,000 moa estimated previously for all species and both islands combined, based on fossil abundance and ecological modelling (Holdaway, Jacomb, 2000).

4.3.5 Phylogeography

As outlined in the introduction to this chapter, a study of the larger phylogeographic patterns was not among the main objectives in this project and will therefore only be discussed briefly. In regards to the central themes in this thesis, the most important information that can be derived from the phylogenetic trees, is arguably that three of the four species exhibited solid evidence of population structuring on the South Island. This finding implies that one of the main assumptions underlying the population size calculation of *Dinornis robustus* in the Gemmell et al. (2004) study was not fulfilled, questioning the accuracy of this estimate, as well as the accompanying interpretation of a pre-human population decline in moa.

Despite being flightless, moa were large and highly mobile animals, so significant barriers in the landscape must have been restricting gene flow to generate the observed patterns. Two of these patterns are discussed here:

1) North Canterbury moa and gene flow along the east coast of the South Island

The presented trees clearly rejected the presence of panmictic moa populations in the South Island. Although the effects of genetic differentiation in time and space can be difficult to disentangle, the few radiocarbon dates available from non-Canterbury material (radiocarbon age information is

embedded in the individual sequence names on the respective trees) showed that the structure was not an effect of time: contemporary individuals were found within separate clades. The Holocene moa did not represent a homogenous gene pool (within each species, respectively), and the sampled North Canterbury individuals belonged to a subpopulation rather than displaying panmixia. Without much wider geographic sample coverage, it is impossible to estimate the size of the area occupied by this genetic subpopulation encompassing North Canterbury, but the extent varied between the four species.

In *Dinornis robustus*, the North Canterbury birds formed a distinct clade with those in South Canterbury and the northwestern South Island (Figure 4.7a). Most individuals from the northwest were, however, embedded in a highly distinctive separate clade (D), which suggested that gene flow was limited between the East and West of the island, because of the Southern Alps. The genetic relationships between North and South Canterbury were strong (there being identical sequences), which suggested that there was a more or less free migration across the Canterbury Plains.

In the phylogenetic trees representing *Euryapteryx curtus* and *Pachyornis elephantopus* respectively, North and South Canterbury individuals were distinctively separated, indicating that gene flow across the Canterbury Plains and along the East Coast was absent during the Holocene. Considering the areas' low relief, only the rivers seem likely to have constituted effective migratory barriers for these species. The Plains are intersected by several major river systems draining the mountains and foothills, and the post-glacial outwash fans may have been sufficiently intense in the early Holocene Canterbury environment to inhibit movement (Bunce et al., 2009; Wilson, 1985). It seems that individuals (females, at least) of *Euryapteryx* and *Pachyornis* were less successful in crossing big rivers than the larger and longer-legged *Dinornis*. The same apparently obstructed gene flow along the East coast has been observed in the genetic signal of the upland moa (*Megalapteryx didinus*) (Bunce et al., 2009).

The limited genetic variability in *Emeus crassus*, meant that it was not possible to detect any population structure. Identical sequences were found in North Canterbury and Southland, which might suggest panmixia throughout its distribution. Because of the low genetic diversity, it is, however, questionable whether geographic population structuring could have been detected, even if it was present. Bunce et al. (2009) suggested that this species had experienced a major bottleneck during the Otiran glaciations, followed by a more recent post-glacial expansion. Nothing in the data presented here suggest otherwise.

2) Central Otago and the northwestern South Island as glacial refugia

Whereas the North Canterbury area was not particularly well defined genetically, two geographic regions; Central Otago and northwestern area, seem to have supported highly divergent and distinct mitochondrial lineages in South Island moa. These are the Northwest area and Central Otago respectively.

Of the four moa species investigated in this project, the northwestern South Island is represented by fossil of *Dinornis* and *Euryapteryx*. For both, the region contained some interesting genetic signatures. Four well-defined lineages can be recognized in *Dinornis*; three of these were present in the northwest. This suggests that the region may have served as a glacial refugium from which *Dinornis* re-dispersed and colonized other parts of the South Island. Two sequences, representing fossils of Pleistocene age, from Honeycomb Hill Cave near Karamea constituted one of the basal lineages in *Dinornis robustus*, with 6-7 unique mutations. Although the two bones have not been radiocarbon dated, 13 other moa bones from Honeycomb Hill Cave have, and they are all 10,000-20,000 years old (Anderson, 1989; Worthy, 1987). The presence of such “old” and highly divergent lineage further supported the concept of an “out of northwest” post-glacial dispersal. Extensive unglaciated landscapes existed in this region during the last glacial maximum and previous molecular work has already suggested this area as a Pleistocene glacial refuge for plants and animals (reviewed in Wallis, Trewick, 2009). Also the *Euryapteryx curtus* tree showed a highly distinct genetic lineage in this area (Figure 4.7c), but the molecular coverage was too limited (only three northwestern sequences) to further pursue an “out of northwest” model. Additionally, Bunce et al. (2009) recognized four different geographically separate genetic clades within the upland moa (*Megalapteryx didiformes*) including a distinct northwestern clade.

In total, clear genetic structuring existed in the data from northwestern South Island, although it was not possible to identify whether this structuring was due to spatial or temporal isolation, or simply mirroring sympatry of old, divergent maternal lineages. It seems evident that understanding the population history in this region is fundamental to understanding Pleistocene-Holocene distribution and dispersal of moa in the South Island.

The Otago area also appeared to have played a central role in the patterns of intraspecific genetic diversity in the Holocene moa. In *Dinornis* and *Pachyornis* and *Euryapteryx* the Central Otago individuals formed separate clades, although the geographic extent of these gene pool was difficult to compare between species because of differences in sample coverage. In *Dinornis*, the two Central Otago individuals formed a basal divergent clade, clearly genetically separated from the South Canterbury individuals. The same pattern was evident in *Pachyornis* but with a close genetic relationship with Southland and the single Stewart Island individual, which was probably brought to the island by early Polynesians (R.N. Holdaway, pers. comm.). In *Euryapteryx*, on the other hand, the

single Central Otago individual grouped strongly with the South Canterbury individuals. Evidence is limited, but this species may have had difficulties crossing the rivers on the Canterbury Plains, while maintaining gene flow up the rivers, thereby connecting Central Otago and the East Coast; in contrast to the other species.

Based on the observation of basal mtDNA lineages in *Megalapteryx* individuals from Otago, Bunce et al. (2009) suggested that the area would have offered upland habitat that was little affected by glaciations and could hence have served as a refugium. This hypothesis has been supported by other molecular studies and biogeographical studies, confirming the area as a biodiversity hotspot (Heads, Patrick, 2003; Wallis, Trewick, 2009).

4.4 Conclusions

In brief, the mtDNA CR data presented here provide a solid basis for understanding the genetic diversity in the four sympatric moa populations. To investigate even the most basic descriptive genetics in this framework is a unique opportunity, and the four haplotype networks are the “crown jewels” in this context. The genetic signals discussed in this chapter appear to reflect stable moa populations, which is supported by the rich North Canterbury fossil record and the temporal deposition patterns (Figure 4.1). In other words; neither the control rRegion data nor the radiocarbon chronology carry signals that could suggest a Holocene population decline in moa, as has been hypothesised previously (Gemmell et al., 2004). Rather, the data indicate that the *Dinornis robustus* population increased in size throughout the early and mid Holocene. The Eastern moa (*Emeus crassus*) displayed low levels of genetic diversity, but this was prevailing throughout the entire sampled period, so it cannot reflect a late Holocene population bottleneck. More likely, it is the result of a demographic bottleneck facilitated by the glaciation. Finally, there was clear evidence of geographic structuring, which excludes the possibility of panmictic moa populations on the South Island.

4.5 References

- Akaike H (1974) New look at statistical model identification. *Transactions on Automatic Control* **19**, 716-723.
- Anderson A (1989) *Prodigious Birds* Cambridge University Press, Cambridge, England.

- Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA (2005) Serial SimCoal: A population genetics model for data from multiple populations and points in time. *Bioinformatics* **21**, 1733-1734.
- Baker AJ, Daugherty CH, Colbourne R, McLennan JL (1995) Flightless brown kiwis of New Zealand possess extremely subdivided population structure and cryptic species like small mammals. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 8254-8258.
- Baker AJ, Huynen LJ, Haddrath O, Millar CD, Lambert DM (2005) Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: The giant moas of New Zealand. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8257-8262.
- Barnett R, Shapiro B, Barnes I, et al. (2009) Phylogeography of lions (*Panthera leo* ssp.) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. *Molecular Ecology* **18**, 1668-1677.
- Bouckaert RR (2010) DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* **26**, 1372-1373.
- Bowie RC, Fjeldsa J, Hackett SJ, Crowe TM (2004) Molecular evolution in space and through time: mtDNA phylogeography of the Olive Sunbird (*Nectarinia olivacea/obscura*) throughout continental Africa. *Molecular Phylogenetics and Evolution* **33**, 56-74.
- Bramanti B, Thomas MG, Haak W, et al. (2009) Genetic discontinuity between local hunter-gatherers and central europe's first farmers. *Science* **326**, 137-140.
- Bunce M, Szulkin M, Lerner HRL, et al. (2005) Ancient DNA provides new insights into the evolutionary history of New Zealand's extinct giant eagle. *Plos Biology* **3**, 44-46.
- Bunce M, Worthy TH, Ford T, et al. (2003) Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425**, 172-175.
- Bunce M, Worthy TH, Phillips MJ, et al. (2009) The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20646-20651.
- Cooper A, Cooper RA (1995) The Oligocene Bottleneck and New-Zealand Biota - Genetic Record of a Past Environmental Crisis. *Proceedings of the Royal Society of London Series B-Biological Sciences* **261**, 293-302.
- Cooper A, Lalueza-Fox C, Anderson S, et al. (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704-707.
- Cooper A, Mourer-Chauvire C, Chambers GK, et al. (1992) Independent Origins of New Zealand Moas and Kiwis. *Proceedings of the National Academy of Sciences* **89**, 8741-8744.
- Cornuet JM, Santos F, Beaumont MA, et al. (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**, 2713-2719.
- de Bruyn M, Hall BL, Chauke LF, et al. (2009) Rapid Response of a Marine Mammal Species to Holocene Climate and Habitat Change. *Plos Genetics* **5**.
- Debruyne R, Chu G, King CE, et al. (2008) Out of America: Ancient DNA Evidence for a New World Origin of Late Quaternary Woolly Mammoths. *Current Biology* **18**, 1320-1326.
- Debruyne R, Poinar HN (2009) Time Dependency of Molecular Rates in Ancient DNA Data Sets, A Sampling Artifact? *Systematic Biology* **58**, 348-359.
- Depaulis F, Orlando L, Hanni C (2009) Using Classical Population Genetics Tools with Heterochroneous Data: Time Matters! *PLoS ONE* **4**.
- Dickerson BR, Ream RR, Vignieri SN, Bentzen P (2010) Population Structure as Revealed by mtDNA and Microsatellites in Northern Fur Seals, *Callorhinus ursinus*, throughout Their Range. *PLoS ONE* **5**.
- Drummond A, Ho SYW, Rawlence NJ, Rambaut A (2007) A rough guide to BEAST 1.4.

- Drummond AJ, Ashton B, Buxton S, et al. (2010) Geneious. Available from www.geneious.com.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**.
- Frankham R (1995) Effective population size / adult population size ratios in wildlife - a review. *Genetical Research* **66**, 95-107.
- Gemmell NJ, Schwartz MK, Robertson BC (2004) Moa were many. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, S430-S432.
- Gilbert MTP, Drautz DI, Lesk AM, et al. (2008) Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 8327-8332.
- Goodman SN (1999) Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine* **130**, 1005-1013.
- Heads M, Patrick B (2003) Biogeography. In: *The Natural History of Southern New Zealand* (eds. Darby J, Fordyce RE, Mark A, Probert K, Townsend C). University of Otago Press, Dunedin, New Zealand.
- Hedrick PW (2000) *Genetics of Populations* Jones and Bartlett Publishers, Inc., London.
- Higham T, Anderson A, Jacomb C (1999) Dating the first New Zealanders: The chronology of Wairau Bar. *Antiquity* **73**, 420-427.
- Ho SYW, Gilbert MTP (2010) Ancient mitogenomics. *Mitochondrion* **10**, 1-11.
- Hofreiter M, Muenzel S, Conard NJ, et al. (2007) Sudden replacement of cave bear mitochondrial DNA in the late Pleistocene. *Current Biology* **17**, R122-R123.
- Hofreiter M, Serre D, Poinar HN, Kuch M, Paabo S (2001) Ancient DNA. *Nature Reviews Genetics* **2**, 353-359.
- Holdaway RN, Jacomb C (2000) Rapid extinction of the moas (Aves : Dinorinithiformes): Model, test, and implications. *Science* **287**, 2250-2254.
- Holdaway RN, Worthy TH (1997) A reappraisal of the late quaternary fossil vertebrates of Pyramid Valley Swamp, North Canterbury, New Zealand. *New Zealand Journal of Zoology* **24**, 69-121.
- Huynen L, Millar CD, Scofield RP, Lambert DM (2003) Nuclear DNA sequences detect species limits in ancient moa. *Nature* **425**, 175-178.
- Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725-&.
- Lambert DM, Ritchie PA, Millar CD, et al. (2002) Rates of evolution in ancient DNA from Adelie penguins. *Science* **295**, 2270-2273.
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.
- Mace GM, Lande R (1991) Assessing extinction threats - toward a reevaluation of IUCN threatened species categories. *Conservation Biology* **5**, 148-157.
- Magoulas A, Castilho R, Caetano S, Marcato S, Patarnello T (2006) Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). *Molecular Phylogenetics and Evolution* **39**, 734-746.
- Mirol PM, Routtu J, Hoikkala A, Butlin RK (2008) Signals of demographic expansion in *Drosophila virilis*. *BMC Evolutionary Biology* **8**.
- Navascues M, Depaulis F, Emerson BC (2010) Combining contemporary and ancient DNA in population genetic and phylogeographical studies. *Molecular Ecology Resources* **doi: 10.1111/j.1755-0998.2010.02895.x**.
- Nevado B, Koblmüller S, Sturmbauer C, et al. (2009) Complete mitochondrial DNA replacement in a Lake Tanganyika cichlid fish. *Molecular Ecology* **18**, 4240-4255.

- Nunney L (1991) The influence of age structure and fecundity on effective population size. *Proceedings of the Royal Society of London Series B-Biological Sciences* **246**, 71-76.
- Nunney L (2000) The limits to knowledge in conservation genetics - The value of effective population size. *Evolutionary Biology, Vol 32* **32**, 179-194.
- Nunney L, Campbell KA (1993) Assessing minimum viable population size - demography meets population genetics. *Trends in Ecology & Evolution* **8**, 234-239.
- Pergams ORW, Barnes WM, Nyberg D (2003) Rapid change in mouse mitochondrial DNA. *Nature* **423**, 397-397.
- Phillips MJ, Gibb GC, Crimp EA, Penny D (2010) Tinamous and Moa Flock Together: Mitochondrial Genome Sequence Analysis Reveals Independent Losses of Flight among Ratites. *Systematic Biology* **59**, 90-107.
- Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.
- Rambaut A, Drummond A (2007) Tracer. Available from <http://beast.bio.ed.ac.uk/Tracer>
- Rambaut A, Ho SYW, Drummond AJ, Shapiro B (2009) Accommodating the Effect of Ancient DNA Damage on Inferences of Demographic Histories. *Molecular Biology and Evolution* **26**, 245-248.
- Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**, 552-569.
- Roman J, Palumbi SR (2003) Whales before whaling in the North Atlantic. *Science* **301**, 508-510.
- Seabrook-Davison M, Huynen L, Lambert DM, Brunton DH (2009) Ancient DNA Resolves Identity and Phylogeny of New Zealand's Extinct and Living Quail (*Coturnix* sp.). *PLoS ONE* **4**.
- Shapiro B, Drummond AJ, Rambaut A, et al. (2004) Rise and fall of the Beringian steppe bison. *Science* **306**, 1561-1565.
- Shapiro B, Sibthorpe D, Rambaut A, et al. (2002) Flight of the dodo. *Science* **295**, 1683-1683.
- Shepherd LD, Lambert DM (2008) Ancient DNA and conservation: lessons from the endangered kiwi of New Zealand. *Molecular Ecology* **17**, 2174-2184.
- Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555-562.
- Stiller M, Baryshnikov G, Bocherens H, et al. (2010) Withering away - 25,000 years of genetic decline preceded cave bear extinction. *Molecular Biology and Evolution* **27**, 975-978.
- Strimmer K, Pybus OG (2001) Exploring the demographic history of DNA sequences using the generalized skyline plot. *Molecular Biology and Evolution* **18**, 2298-2305.
- Turvey ST, Green OR, Holdaway RN (2005) Cortical growth marks reveal extended juvenile development in New Zealand moa. *Nature* **435**, 940-943.
- Turvey ST, Holdaway RN (2005) Postnatal ontogeny, population structure, and extinction of the giant moa *Dinornis*. *Journal of Morphology* **265**, 70-86.
- Wallis GP, Trewick SA (2009) New Zealand phylogeography: evolution on a small continent. *Molecular Ecology* **18**, 3548-3580.
- Willerslev E, Cooper A (2005) Ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **272**, 3-16.
- Wilmshurst JM, Anderson AJ, Higham TFG, Worthy TH (2008) Dating the late prehistoric dispersal of polynesians to New Zealand using the commensal Pacific rat. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 7676-7680.
- Wilson DD (1985) Erosional and deposition trends in rivers in the Canterbury Plains, New Zealand. *Journal of hydrologi, New Zealand* **24**, 32-44.
- Worthy T (1987) The subfossil fauna of Honeycomb Hill Cave.

- Worthy TH (2002) The youngest giant: discovery and significance of the remains of a giant moa (*Dinornis giganteus*) near Turangi, in central North Island, New Zealand. *Journal of the Royal Society of New Zealand* **32**, 183-187.
- Worthy TH, Holdaway RN (1996) Quaternary fossil faunas, overlapping taphonomies, and palaeofaunal reconstruction in north Canterbury, South Island, New Zealand. *Journal of the Royal Society of New Zealand* **26**, 275-361.
- Worthy TH, Holdaway RN (2002) *The lost world of the Moa* Canterbury University Press, Christchurch.

Chapter 5: A microsatellite study of the North Canterbury moa: population genetics in a temporal dimension

A microsatellite study of the North Canterbury moa: population genetics in a temporal dimension

5.1 Introduction

This chapter represents the first population study to be based on microsatellite data from ancient substrates. The analyses presented here take aDNA research into ‘uncharted territory’ by offering high resolution genetic insights into four extinct populations of moa. Whereas the technology and criteria for generating aDNA microsatellite data were established in Chapter 3, the present chapter displays the outcome of the further analyses.

There were several strong incentives for engaging in microsatellite work despite well known challenges in amplifying nuclear DNA from ancient substrates (e.g., Willerslev, Cooper, 2005). In the earliest stages of this project, it became clear that the biomolecule preservation in the North Canterbury fossils is remarkable (see Chapter 2). It seemed obvious that if a microsatellite population study based on ancient DNA templates was possible, these moa bones had the preservation to yield a reliable dataset. Also, if indeed the data could be obtained, the paleontological setting with hundreds of fossils preserved in close spatial and temporal proximity, offered a framework for applying the allele frequency based analyses that characterise microsatellite population studies. Finally, a microsatellite dataset would offer an independent and novel proxy to study moa genetics with some obvious benefits compared to the mtDNA described in Chapter 4. The mutation rate in microsatellites is higher than in mtDNA. Typically it is estimated to between 10^{-7} to 10^{-3} per locus per generation (summarised in Buschiazzi, Gemmell, 2006), whereas the mtDNA CR mutation rate in moa was estimated to 8.7×10^{-8} per site per year (Bunce et al., 2009). Moreover, a variable microsatellite locus can display numerous alleles, in contrast to the maximum of four states in a single nucleotide polymorphism, so microsatellite polymorphisms will often prove more informative at the population level compared to mtDNA. Analyses of such two datasets in concert (mtDNA and microsatellites) would provide insights at two different evolutionary time-scales and also reflect two different evolutionary processes, since mtDNA is inherited maternally, whereas the nuclear microsatellites are passed on by both sexes.

The generation of microsatellite data from aDNA templates represents a significant technical advance in molecular research, but the data offer considerable analytical challenges. With heterochronous data, the observed genetic variation is not restricted to spatial patterns but can vary in two dimensions (time and space). This means that classical assumptions in population genetic analyses, such as random mating and Hardy-Weinberg proportions, are at least conceptually violated. This implies that the results must to be interpreted with caution and that analytical tools are not (yet) available to take full

advantage of the information embedded in such ‘time stamped’ microsatellite data. Regardless, this chapter is perceived as a solid exploration of the first microsatellite data collected from extinct species, displaying the potential and the limitations, while setting the scene for future research. An array of classical microsatellite analyses have been carried out below, but with interpretational caveats to make them relevant for data that cover thousands of years.

With four different microsatellite datasets at hand, along with four mtDNA datasets, accompanied by gender and maturity information, fossil site and radiocarbon age for each individual (Digital Appendix 2), there are numerous potential signals and correlations to investigate. This chapter will specifically address the following six areas:

1) Basic descriptive statistics, tests for Hardy-Weinberg (HW) proportions and linkage

This part includes estimators and tests that are routinely carried out as a basic description of the data in microsatellite studies. Measures of genetic diversity provide an overview of the variability in each of the four moa taxa, comparable to the basic mtDNA analyses in Chapter 4. It can therefore be tested with a high level of accuracy whether the nuclear gene pools mirror the mtDNA result of low genetic diversity in *Emeus crassus* compared to the other three species (see Chapter 4). Tests for HW-proportions are also highly relevant. They will first of all indicate whether the data is compromised by allelic dropout (see Chapter 3), but they can also provide information on the genetic structure. If the data reflect several strongly separated gene pools with deviating allele frequencies, a deficiency of heterozygous individuals is expected (the Wahlund effect) compared to the HW-proportions. However, for this heterochronous data, we cannot expect HW-proportions by default despite the spatial proximity of the sites. This fact represents one of the main caveats in the microsatellite analyses. Even if a North Canterbury moa population was characterised by random mating and large homogenous gene pool, the samples have not been obtained at the same point in time, and any deviations from HW-proportions could therefore be a temporal effect, rather than an effect of allelic dropout or spatial structuring. On the other hand, if indeed the data show HW-proportions, it would support the argument of very limited genetic drift during the late Holocene and no major loss of alleles due to a population decline. The same arguments apply for tests of linkage disequilibrium (LD) within each dataset. If certain alleles were introduced by mutations or genetic influx at a particular point in time, it could cause signals of LD, because the older genotypes have not been able to combine with the new alleles. If no signal of LD can be detected, it indicates that allele frequencies were stable through time.

2) Genetic diversity through time

Plotting individual observed heterozygosity (H_O) against radiocarbon age should provide unique empirical insight into temporal aspects of microsatellite variability. The analysis will show the span of individual H_O within the population, but more interestingly, it should also reveal if significant changes in genetic diversity happened over the >4000 years covered by the samples. For neutral loci such as microsatellites, genetic drift eliminates heterozygosity in a finite population with a constant rate of $1/(2N_e)$ per generation (Hedrick, 2000; Kimura, Crow, 1964), implying that the loss of genetic diversity is high in small populations. Hence, if moa went through a major Holocene decline prior to the arrival of humans, as hypothesised by Gemmell et al. (2004), the expectation would be a corresponding decline in individual heterozygosity through time – depending on the severity and extent of this proposed bottleneck. Given that no decline-signal could be recovered based on the mtDNA (Chapter 4), the microsatellite data are expected to display no obvious loss of genetic diversity.

3) Genetic structure

Because of the close proximity of the North Canterbury study sites (Chapter 1, Section 3), space was not expected to influence genetic structure in the data. As for the mtDNA data, time was more likely to have had an effect (i.e., altered the allele frequencies), either through gradual background genetic drift, or through abrupt demographic events causing alleles to disappear or new alleles to appear by a sudden genetic influx. Several methods can be applied to detect any change in allele frequencies over time. These include methods that are normally used to define population boundaries, as well as methods for detecting isolation by distance. A distinct correlation between geographic and genetic distance is expected when gene flow between populations is restricted by distance (Wright, 1943). However, it seems reasonable that the same pattern would apply to time – isolation by time (IBT). As mentioned above, genetic drift is expected to alter allele frequencies at a pace defined by population size, whereas incoming gene flow and new mutations increase the diversity. Hence, individuals separated by thousands of years are expected to be more genetically distant than contemporary individuals. This hypothesis of IBT is tested in this chapter as another method of assessing the level of genetic stability through time. If only subtle background levels of genetic drift have occurred (for example due to large stable populations) then only a weak (or no) correlation between genetic distance and time is expected.

4) Identifying relatives

The high variability in microsatellites provides great resolution power to identify relatives. Several methods and software packages have been developed for these types of analyses (reviewed in Blouin, 2003). With hundreds of moa fossils accumulated in a small area, it is not unlikely that relatives could have been deposited, excavated and sampled. Such finding would potentially provide new insights on the biology of moa, and perhaps support or reject the hypotheses regarding territoriality and dispersal proposed in Chapter 2. The presence of relatives in the same swamp could for example indicate that moa did not disperse very far from the area where they hatched. From a technical view, it was relevant to investigate whether this cornerstone in microsatellite-based research, namely kinship analyses, applied to time stamped data. In this context, the radiocarbon ages represented excellent means to validate the kinship results, because of the simple fact that closely related individuals must have been contemporary.

5) Detecting migrants

Microsatellite data represent a powerful tool to identify related individuals based on the genetic similarities. Hence, they can also be used to identify individuals with atypical genetic profiles. Migrants (or recent descendants of migrants) can be identified with confidence, if the allele frequencies between source and sink population differ. It was therefore analysed, based on a self-assignment test (Paetkau et al., 2004), whether distinct genetic outliers were present among the sampled moa individuals. This analysis, provided an opportunity to investigate the concordance between nuclear and mtDNA in revealing genetic anomalies. Detection of migrants would provide information on the degree of genetic connectivity between moa subpopulations. The mtDNA analyses (Chapter 4) showed that the South Island moa populations were structured, so if any migrants from different population have been preserved in North Canterbury, and sampled for this project, the microsatellite data should hopefully reveal this.

6) Sex biased dispersal

Sex biased dispersal is a common trait in many animals. When present in mammals, males are commonly found to be the dispersing sex, whereas gene flow is most often facilitated by females in birds (Clarke et al., 1997; Dobson, 1982; Greenwood, 1980; Prugnolle, de Meeus, 2002). Given that moa were atypical birds in many respects, with extreme reverse sexual dimorphism (Bunce et al., 2003; Huynen et al., 2003) and possibly female territoriality (Chapter 2) and male incubation (Birchard, Deeming, 2009), sex roles could also have been reversed in terms of dispersal. The female

territorial behaviour hypothesised in Chapter 2, could perhaps suggest that males were dispersing more than females and it is therefore tested here. It is challenging to assess dispersal, migration and gene flow when all genetic samples are from the same area, but several different approaches, based on relatedness values (Queller, Goodnight, 1989) and assignment tests (Favre et al., 1997; Mossman, Waser, 1999), were applied in an attempt to shed light on this matter, as it could potentially provide valuable insights into moa biology.

In summary, the aims of this chapter are to provide new perspectives on moa biology, and at the same time to thoroughly explore these heterochronous microsatellite data to reveal strengths and weaknesses.

5.2 Methods and results

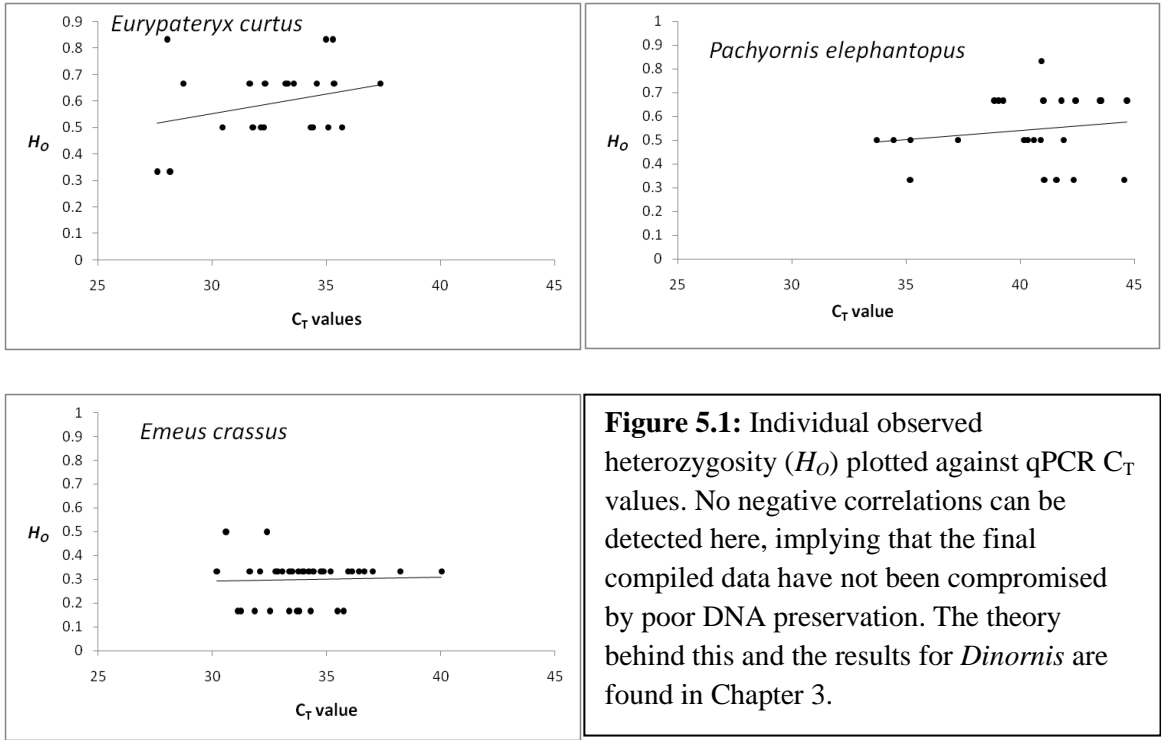
5.2.1 Data generation

All information regarding primer design and PCR conditions can be found in Chapter 3. Note that identical PCR conditions (Table 3.2.2) applied to all four moa species. To generate reliable microsatellite data for *Emeus crassus*, *Pachyornis elephantopus* and *Euryapteryx curtus*, the three criteria established for *Dinornis robustus* were used (see Chapter 3, Section 2). Although the extent of allelic dropout was not assessed in separate comprehensive pilot studies for the three other moa species, there were no indications that dropout rates differed between taxa, so the $\geq 4 \times$ PCR approach for apparent homozygotes was adopted for all species. Exceptions were made, however, in markers without genetic variability, where the recorded allele was accepted after just one positive reaction. Because of the considerable effort, required to generate reliable microsatellite data from aDNA templates, the work on *P. elephantopus*, *E. curtus* and *E. crassus* was focused on DNA extracts from the 217 radiocarbon dated bones – primarily because time-stamped data are far more informative than data with unknown temporal context. After the full microsatellite datasets had been compiled, they were subjected to the same quality-checks as the *Dinornis* data (Chapter 3, Section 2). By plotting qPCR C_T values against individual observed heterozygosity (H_O) (Figure 5.1) it was shown that poor DNA quality was not affecting the recorded H_O as would be expected if allelic dropout was compromising the data (discussed in Chapter 3, Section 2).

It should be noted, that the lengths of some alleles in the data have not yet been fully verified. In some occasions (some taxon-locus combinations: see Digital Appendix 13), the recorded alleles lengths do not fit the expected di-nucleotide repeat motif throughout the entire allelic range – meaning that both odd and even numbers of base pairs occur. This could be due to a mutation (insertion) linked to certain alleles in some species, but preliminary analyses suggest that it is an issue with incorrect, yet

consistent, reading of alleles of particular lengths. Cloning and sequencing are presently being undertaken to fully resolve this issue. Unless where noted, this issue will not affect the analyses and interpretations presented in this chapter as all alleles have been scored consistently across individuals. Importantly; consistent genotyping was ensured through multiple PCR repeats of each problematic individual, which allowed for the establishment of strict allele-call-bins (Digital Appendix 13). The issue does not influence measures of genetic diversity, since a heterozygote is a heterozygote regardless of the allele lengths. Moreover, the applied microsatellite analyses do not incorporate a mutation model (except for the isolation by time (IBT) analysis, as discussed under that section), so the exact allele-lengths are not important *per se*, as long as they have been genotyped consistently.

Figure 5.1



5.2.2 Tests for Hardy-Weinberg (HW) proportions and linkage disequilibrium (LD)

Tests for HW-proportions, LD, and the presence of null alleles were performed as described in Chapter 3 (Section 2). Table 5.1 shows the HW-test results for all 24 taxon-locus combinations, largely confirming congruence between observed and expected heterozygosity. Only two significant deviations were observed and the *Dinornis*-Moa_MA21 deviation was not significant after Bonferroni correction (see Chapter 3, Section 2). A slightly more severe heterozygote deficiency was observed in

Pachyornis-Moa_MA1, proving significant even after Bonferroni correction. As mentioned, HW-proportions are not expected by default in heterochronous data, but the fact that a significant deficiency within this species occurred in a single locus only, could perhaps suggest a problem with this marker. Cloning and sequencing are ongoing to shed light on this matter.

All tests for linkage disequilibrium among loci proved insignificant (P ranging from 0.07 to 1) except for one occasion being an apparent Moa_MS2/Moa_MA1 linkage in *Pachyornis* ($P = 0.0005$). Unless a chromosomal rearrangement has occurred in this species alone, the result is unlikely to represent true physical linkage and is rather an effect of non-random mating (for example caused by heterochronous data). A simple visual inspection of the *Pachyornis* data suggests that the linkage is caused by a 106/106 bp homozygous profile in the Moa_MS2 locus which is always (as in five individuals) accompanied by a 90/90 bp homozygous profile in Moa_MA1 (Digital Appendix 14). Finally, MICROCHECKER ver. 2.2.3 (van Oosterhout et al., 2006) rejected the presence of scoring bias as a result of null alleles, allelic dropout or stuttering in the four datasets.

Table 5.1: Fixation index (F_{IS}) and tests for HW-proportions in the four moa populations. F_{IS} was calculated according to Weir and Cockerham (1984) and the P value for each locus-taxon deviation was assessed with the exact test implemented in GENEPOP ver. 4.0.10 (Rousset, 2008). Overall P -value within each taxon was calculated using Fishers approach for combining probabilities. Red values represent significant deviations ($\alpha = 0.05$). NA represents fixed or nearly fixed loci. Values of expected (H_E) and observed (H_O) heterozygosity for each locus-taxon combination can be found in Digital Appendix 15.

Taxon	n	MS2		MA1		MA21		MA38		MA44		MA46		Overall	
		F_{IS}	P	F_{IS}	P	F_{IS}	P	F_{IS}	P	F_{IS}	P	F_{IS}	P	F_{IS}	P (Fisher)
<i>D. robustus</i>	74	0.027	0.674	-0.010	0.897	0.116	0.040	0.061	0.331	0.080	0.257	0.010	0.238	0.040	0.227
<i>P. elephantopus</i>	30	0.191	0.392	0.287	0.002	-0.150	0.582	-0.018	0.636	-0.036	1	0.046	0.444	0.041	0.118
<i>E. curtus</i>	29	-0.043	0.816	-0.069	1	0.037	0.171	-0.129	0.914	-0.018	1	-0.146	1	-0.079	0.981
<i>E. crassus</i>	55	NA	NA	-0.031	1	-0.035	0.284	-0.089	0.781	NA	NA	NA	NA	-0.048	0.807

5.2.3 Genetic diversity

Measures of genetic diversity were calculated in MSTool ver. 3.1.1 (Park, 2001) and are shown in Table 5.2. The underlying allele scores and allele frequencies can be found in Digital Appendix 14 and 16 respectively. Of the 217 radiocarbon dated fossils, 183 (84.3%) produced a multi-locus microsatellite genotype according to the criteria established in Chapter 3 (Section 2). *Euryapteryx curtus* showed a lower success rate than the other species (65 %) (Table 5.2). This can be ascribed to a higher fraction of juveniles in this species, having a less dense bone matrix and a poorer DNA preservation (discussed in Chapter 3). In addition to the 183 profiles from dated fossils, microsatellite

profiles had been generated from five undated bones during the PCR optimisation protocols, increasing the total number of multi-locus profiles to 188 (Table 5.2).

The level of genetic diversity differed among the four sympatric populations. A certain ascertainment bias is expected however, because these six markers were developed specifically for *Dinornis robustus* (see Chapter 3, Section 2) and secondarily tested on the other species (Goldstein, Pollock, 1997). This means that loci proving mono- or di-morphic in *Dinornis* were excluded regardless of their variability in the other species. Yet, this bias is easily removed by only comparing the three emeids (*Pachyornis*, *Euryapteryx*, *Emeus*). The conclusion remains that *Emeus crassus* displayed very low genetic diversity compared to the other species. In fact, two of the six markers are fixed and two are di-morphic for this species (Digital Appendix 14).

Table 5.2: Measures of genetic diversity across 188 profiled individuals (n) and four taxa. Multi-locus genotyping success rate is given for each species, as well as average number of observed alleles per locus (N_A) number of effective alleles (N_E), and observed (H_O) and expected heterozygosity (H_E) accompanied by the overall fixation index (F) for each taxon.

Taxon	n	Success	N_A	N_E	H_O	H_E	F
<i>D. robustus</i>	74	84.1 %	9.5	3.8	0.687	0.721	0.040
<i>P. elephantopus</i>	30	93.8 %	7.2	3.9	0.561	0.601	0.041
<i>E. curtus</i>	29	65.0 %	8.7	4.6	0.600	0.570	-0.079
<i>E. crassus</i>	55	93.0 %	4.7	2.4	0.305	0.288	-0.048

Figure 5.2a

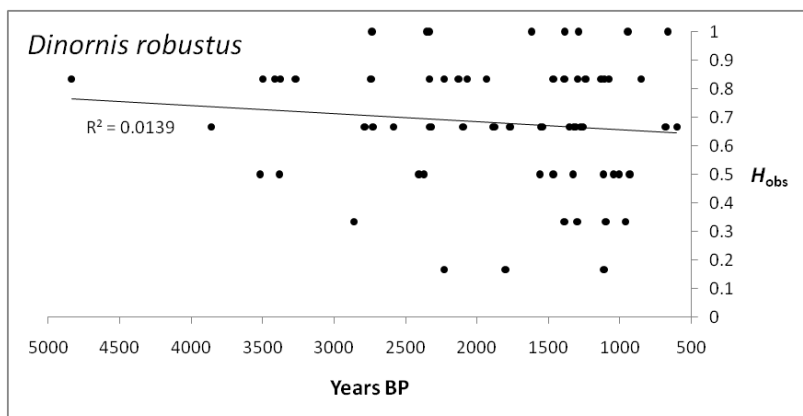


Figure 5.2a-d: Individual observed heterozygosity (H_O) across the six microsatellite loci, plotted against radiocarbon age.

Figure 5.2b

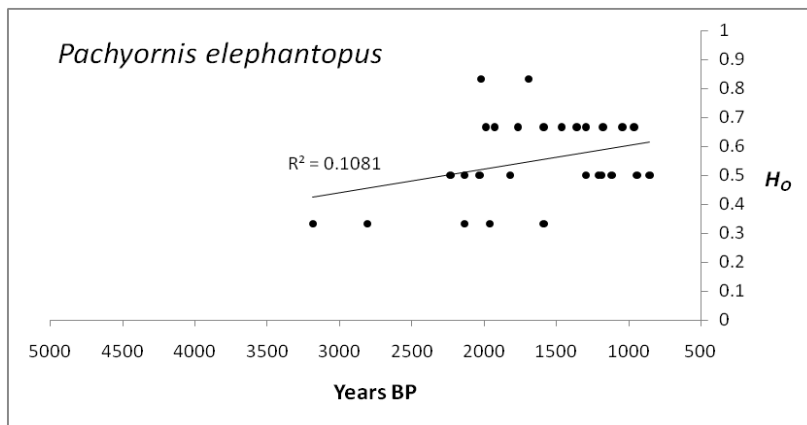


Figure 5.2c

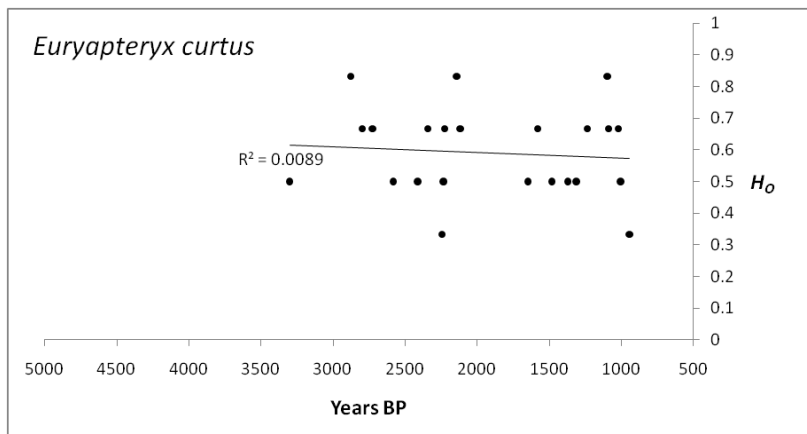
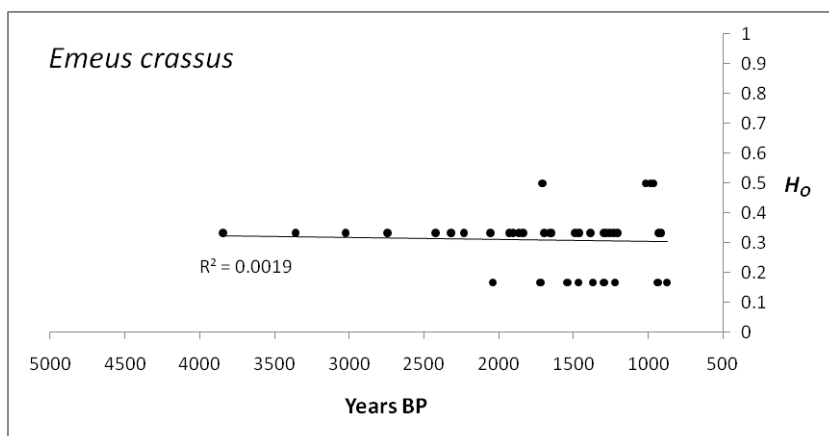


Figure 5.2d



5.2.4 Heterozygosity through time

To investigate the level of genetic diversity through time, individual heterozygosity (H_o = proportion of heterozygous loci) was plotted against radiocarbon age (Figures 5.2a-d). Although large variance, the four plots showed no indications of either loss or gain of genetic variability throughout the sampled time period. The significance of each potential correlation was assessed in an ANOVA regression analysis, yielding P values as follows: *Dinornis robustus*, $P = 0.33$; *Pachyornis elephantopus*, $P = 0.08$; *Euryapteryx curtus*, $P = 0.20$; *Emeus crassus*, $P = 0.77$.

5.2.5 Genetic structure

The programme STRUCTURE ver. 2.3.3 (Falush et al., 2003; Pritchard et al., 2000) represents a powerful Bayesian approach to identify genetic structure in microsatellite data (i.e. distinct groups of individuals). Without using any prior information (sampling locality or radiocarbon age), STRUCTURE clusters individuals into ‘natural populations’ by minimising deviations from HW-proportions and LD, thereby increasing the overall likelihood of the data (Pritchard et al., 2000). Several STRUCTURE analyses were carried out:

First, all individuals across all four species were pooled and analysed. This was done to assess if the data contained any genetic signal in terms of differing allele frequencies. If STRUCTURE was incapable of separating the four species, it would likely be problematic to investigate the more subtle intraspecific patterns. The analysis was run by assuming four, five and six genetic groups respectively ($K = 4, 5, 6$), with each run repeated twice to ensure consistency. Default settings were applied, allowing for genetic admixture and correlated allele frequencies. Both circumstances are somewhat unlikely given that the combined dataset represents four different species. However, since this initial analysis was conducted to illuminate the power of resolution rather than being biologically correct, the same priors were chosen, as would be applied in the subsequent intraspecific assessments. The MCMC was run for 10^6 iterations and a burn-in of 10^4 states. The estimated admixture proportions (Q) are shown in Table 5.3a-c, accompanied by bar plots in Figure 5.3.

At $K = 4$ (assuming four genetic clusters), STRUCTURE identified the four species very distinctively, with Q -values strongly favouring a single genetic unit for each species (Table 5.3a). This was a solid confirmation that the microsatellite data was informative at the basal level (i.e. could distinguish between different species). It should, however, be emphasized that in this particular taxon-combined STRUCTURE analysis, the issue of unknown allele lengths could slightly bias the results. In the hypothetical scenario that identical alleles are present in several species, but have been genotyped consistently at different lengths in each species, the species boundaries will be artificially strengthened.

In the well defined genetic structure of $K = 4$, only one of the 188 genotyped individuals appeared genetically misplaced (Figure 5.3). This was a female *Dinornis robustus* (AV8470) from Pyramid Valley dated to 3376 BP. This individual did not show any unique alleles but simply a rare combination of alleles (Digital Appendix 14).

Figure 5.3: STRUCTURE run of all four species combined

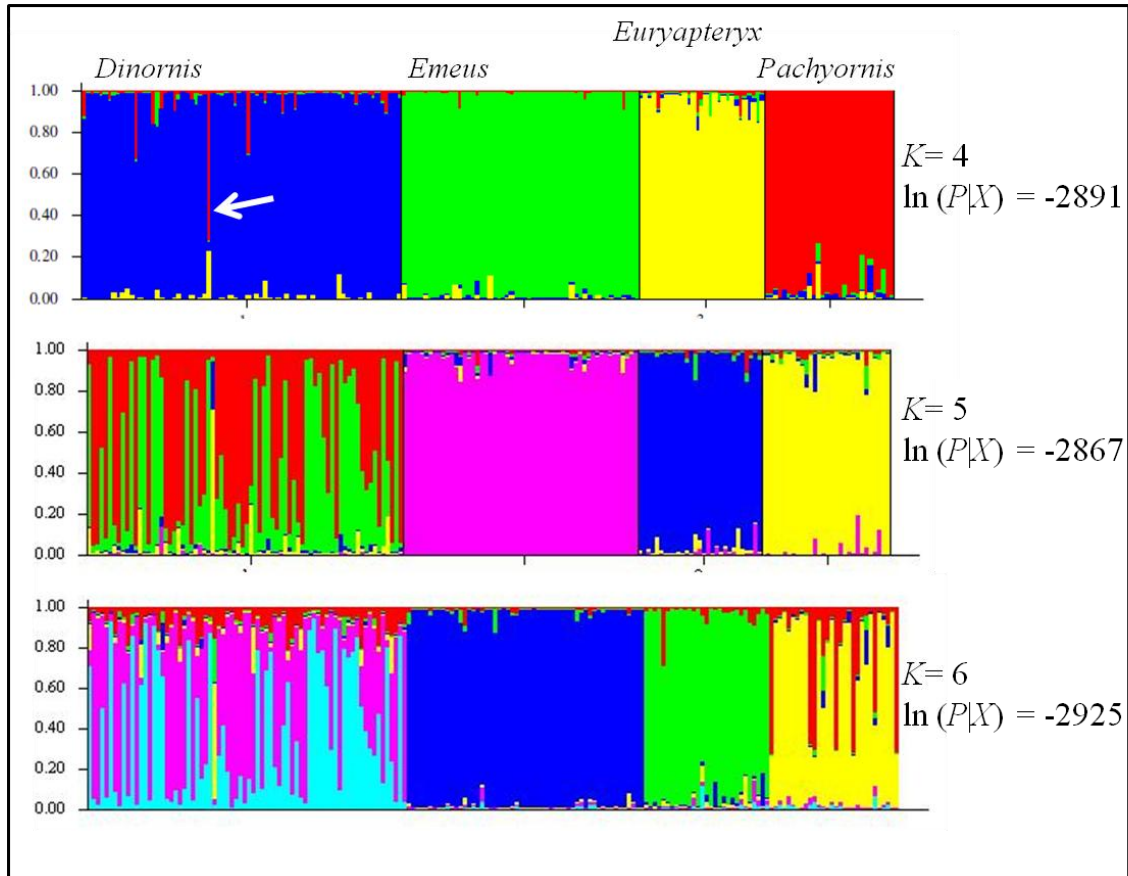


Figure 5.3: Visual output (bar plots) from STRUCTURE runs with combined dataset assuming four, five and six genetic clusters respectively ($K = 4, 5, 6$). Likelihood values (mean from two runs) are shown for each run. Each vertical line represents an individual (188 in total) with colour composition referring to admixture proportions. The bar plot for $K = 4$ confirmed that the genetic structure was defined by taxon. Note a single *Dinornis* individual (AV8470) showing little genetic affiliation with the *Dinornis* gene pool (white arrow). $K=5$ showed highest likelihood value, with *Dinornis* individuals separated in two clusters, supported by relatively high individual admixture proportions (i.e. each vertical line is either red or green), whereas $K=6$ provided a less convincing signal (the structuring within *Pachyornis* is not well defined) and lower overall likelihood.

Table 5.3a: Overall admixture proportions (Q -values) for each species, when assuming four genetic clusters ($K = 4$). The four clusters are distinctively defined by taxon.

	1	2	3	4
<i>D. robustus</i>	0.935	0.040	0.009	0.017
<i>P. elephantopus</i>	0.018	0.944	0.022	0.017
<i>E. curtus</i>	0.016	0.021	0.019	0.944
<i>E. crassus</i>	0.007	0.009	0.970	0.013

Table 5.3b: Overall admixture proportions (Q -values) for each species, when assuming five genetic clusters ($K = 5$). The *Dinornis* gene pool is shared between cluster 1 and 5.

	1	2	3	4	5
<i>D. robustus</i>	0.524	0.036	0.008	0.015	0.418
<i>P. elephantopus</i>	0.015	0.930	0.023	0.017	0.016
<i>E. curtus</i>	0.013	0.020	0.020	0.932	0.014
<i>E. crassus</i>	0.007	0.009	0.965	0.013	0.007

Table 5.3c: Overall admixture proportions (Q -values) for each species, when assuming six genetic clusters ($K = 6$). The *Pachyornis* gene pool is shared between cluster 2 and 6, but as seen in Figure 5.3, this is not supported by high individual admixture proportions.

	1	2	3	4	5	6
<i>D. robustus</i>	0.484	0.031	0.008	0.014	0.394	0.069
<i>P. elephantopus</i>	0.013	0.722	0.021	0.015	0.013	0.215
<i>E. curtus</i>	0.013	0.018	0.021	0.903	0.013	0.032
<i>E. crassus</i>	0.007	0.007	0.958	0.012	0.006	0.009

Interestingly, at $K = 5$, the *Dinornis* gene pool was split up into two clusters (group 1 and 5, Table 5.3b) with the majority of individuals displaying relatively high Q -values (i.e. Q often >0.8 , visualised in Figure 5.3). Moreover, $K = 5$ yielded higher overall likelihood than $K = 4$ (Figure 5.3), supporting this structuring of the *Dinornis* gene pool. However, a closer inspection of the information associated with each fossil revealed that this apparent structure was not defined by either time (radiocarbon age), space (different fossil sites), or by the two distinct mtDNA haplogroups presented in Chapter 4. In fact, by examining the raw data, it was difficult to recognise any consistent genetic differences justifying the separation into two *Dinornis* groups. At $K = 6$ the *Pachyornis* gene pool was separated into two clusters, but since this was not accompanied by high individual affiliation to a single cluster (visualised on Figure 5.3), the structure cannot be considered ‘real’. This interpretation was also supported by a smaller likelihood values for $K = 6$ (Figure 5.3).

To investigate intraspecific genetic structure, microsatellite data from each taxon was run separately at $K = 1, 2, 3$, with the same settings as outlined above. No signal could be detected in any of the four species from these analyses. The output for *Dinornis robustus* at $K = 2$ is shown below (Figure 5.4), reflecting a completely admixed gene pool (bar plots of $K = 2$ for the three other species are shown in Digital Appendix 17). This is contradicting the separation of *Dinornis* into two clusters, observed at $K = 5$ (Figure 5.3). However, given that neither time nor space were identified as responsible for this apparent clustering at $K = 5$, it seems likely to be an artefact of pooling the data across species. The separation of *Dinornis* may have happened because this species exhibits higher genetic variability than the other species.

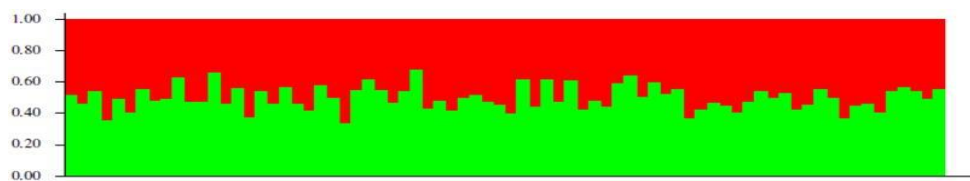


Figure 5.4: STRUCTURE run of multi-locus data from 74 *Dinornis robustus* individuals, assuming two genetic clusters ($K=2$). The signal shows a homogenous gene pool with each individual equally represented in both clusters.

Table 5.4

	1	2
602 BP	0.500	0.500
666 BP	0.503	0.497
680 BP	0.500	0.500
851 BP	0.500	0.500
928 BP	0.499	0.501
931 BP	0.500	0.500
944 BP	0.500	0.500
962 BP	0.500	0.500
1004 BP	0.499	0.501
1043 BP	0.500	0.500
3270 BP	0.501	0.499
3376 BP	0.503	0.497
3380 BP	0.501	0.499
3383 BP	0.499	0.501
3416 BP	0.500	0.500
3497 BP	0.500	0.500
3516 BP	0.501	0.499
3858 BP	0.499	0.501
4126 BP	0.501	0.499
4837 BP	0.502	0.498

Table 5.4: Individual admixture proportions (Q -values) for the ten youngest (602-1043 BP) and ten oldest (3270-4837 BP) *Dinornis robustus* individuals, in a STRUCTURE run assuming two genetic clusters ($K=2$). The 50/50 affiliation reflects a homogenous gene pool despite the 2200 BP gap between the old and the young batch of individuals.

In the final analysis with STRUCTURE, data from the 10 oldest and 10 youngest individuals were pooled to investigate if any shifts in allele frequencies had happened during the intervening time period. For *Dinornis robustus*, the two temporal groups covered 4837-3270 BP and 602-1043 BP respectively, leaving a 2200 year gap for genetic changes to have occurred. The result showed no traces of genetic structure in any of the species, suggesting genetic stability through time. Table 5.4 shows the results from the “old vs young” run of *Dinornis* at $K = 2$. Analyses of the other three species yielded similar results (all four bar plots for “old vs young” at $K = 2$ are shown in Digital Appendix 17).

In addition to the STRUCTURE analyses, a more classical approach was applied to investigate genetic structure. The fixation index (F_{ST}) expresses the proportion of genetic variation in a dataset that can be attributed to differentiation among groups of individuals (Hedrick, 2000). In many population genetic contexts, F_{ST} is used as a measure of genetic differentiation between subpopulation. Since geographic differentiation was not a factor in this study, the ‘subpopulations’ were here constructed by pooling individuals into radiocarbon age groups, to investigate genetic differentiation in time. Only the *Dinornis* data were investigated with this method, as they contained the highest level of genetic variation and most radiocarbon dates, implying that any temporal changes should be easier to uncover with this dataset. Moreover, *Dinornis robustus* is the species for which a major Holocene decline has been suggested (Gemmell et al. 2004), which is why it is most relevant to assess the degree of genetic stability in this species.

Individuals were separated into four age groups, simply representing different millennia (Table 5.5). F_{ST} values between all pairs of groups were calculated and the significance of the observed differentiation was assessed by a permutation test implemented in FSTAT ver. 2.9.3 (Goudet, 2001).

Only small fractions of the genetic variation in the overall dataset could be ascribed to differentiation between age groups, and none of the pairwise F_{ST} values proved significantly higher than values representing permuted data (Table 5.5). This F_{ST} analysis supported the more advanced Bayesian approach above, confirming a complete absence of genetic structure from the *Dinornis* gene pool, despite having been sampled across >4000 years.

Table 5.5: Pairwise F_{ST} values (lower half) between the four different age groups in *Dinornis robustus*, and P -values (upper half) calculated as the fraction of 6000 randomizations resulting in same or higher F_{ST} values than the observed.

	n	<1000 BP	1000-2000 BP	2001-3000 BP	>3000 BP
<1000 BP	8	—	0.26	0.23	0.22
1000-2000 BP	36	0.004	—	0.36	0.44
2001-3000 BP	20	0.013	0.001	—	0.10
>3000 BP	10	0.028	-0.003*	0.012	—

* F_{ST} values normally range from 0 to 1, but slightly negative values can emerge when estimates are weighted for differences in variance [here according to the Weir and Cockerham (1984) estimator implemented in FSTAT].

5.2.6 Isolation by time

No sudden shifts in allele frequencies could be detected in the analyses above, but that does not necessarily imply that the gene pools were static during the investigated time frame. Low background levels of genetic drift or changes in incoming gene flow can generate subtle gradual genetic modifications in a population over time, but more sensitive methods are required to uncover such signal. A Mantel test is commonly applied to assess correlations between genetic and geographical distances in population genetic analyses, but this method should be equally suitable to detect isolation by time (IBT). Matrixes of temporal and genetic distances between all pairs of individuals (in each species separately) were generated and tested for a positive correlation by using a Mantel test implemented in GenAlEx ver. 6 (Peakall, Smouse, 2006). The temporal distances were recorded as the time (in years) separating two individuals. GenAlEx calculates each pairwise genetic distance as $GD = (S1 - S2)^2$, where $S1$ is the size of allele 1 and $S2$ the size of allele 2, and then summed across all loci. The P -value for each correlation was assessed with 1000 randomisations.

This genetic distance, assumes a stepwise mutation model, where individuals with larger deviations in allele sizes are believed to be more distantly related. This can be seen as an alternative to the F_{ST} measure, which failed to reveal a pattern in the analysis above. The F_{ST} measure is based solely on “raw” allele frequencies, and does not discriminate between relative differences in allele lengths (i.e. assumes an infinite allele mutation model). Since the GD distance implemented in GenAlEx uses allele lengths, it also represents an analysis that is potentially compromised by incorrect allele lengths. As mentioned, alleles have been genotyped consistently, so this issue is unlikely to induce a bias – in particular because the magnitude of a potential mis-scoring would be one base-pair only (i.e. an even numbered allele length has been assigned an odd number or vice versa) with little effect on GD.

The plots showed no pattern of isolation by time in any of the four gene pools (Figure 5.5a-d) with P -values of 0.48, 0.24, 0.37 and 0.35 for *D. robustus*, *P. elephantopus*, *E. curtus* and *E. crassus* respectively.

Figure 5.5a

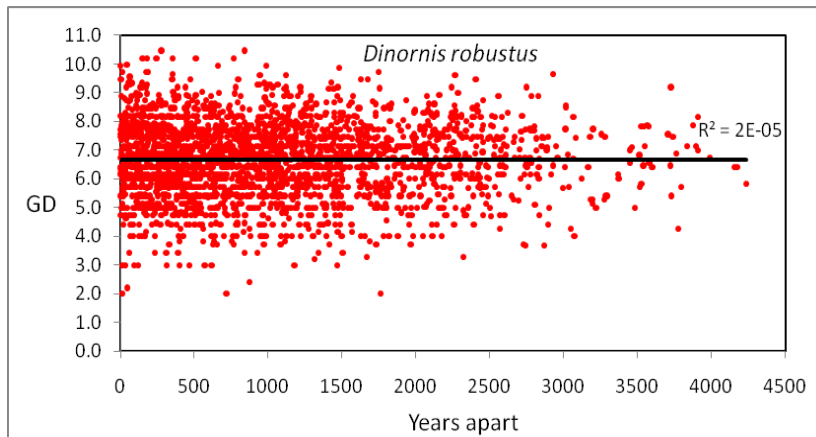


Figure 5.5a-d:

Investigations of isolation by time. The pairwise genetic distances (GD) between all individuals were not correlated with temporal distance (years apart) in any of the four moa species.

Figure 5.5b

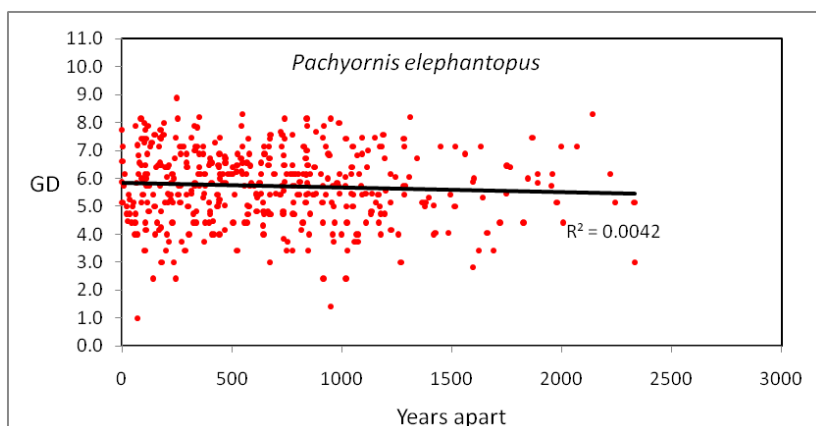


Figure 5.5c

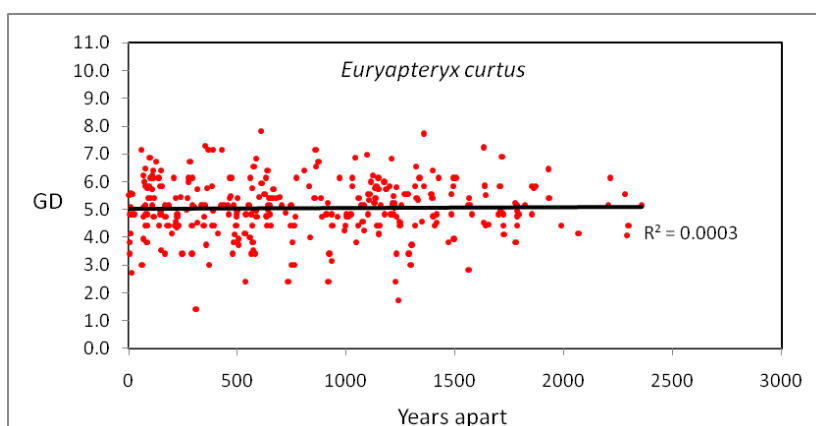
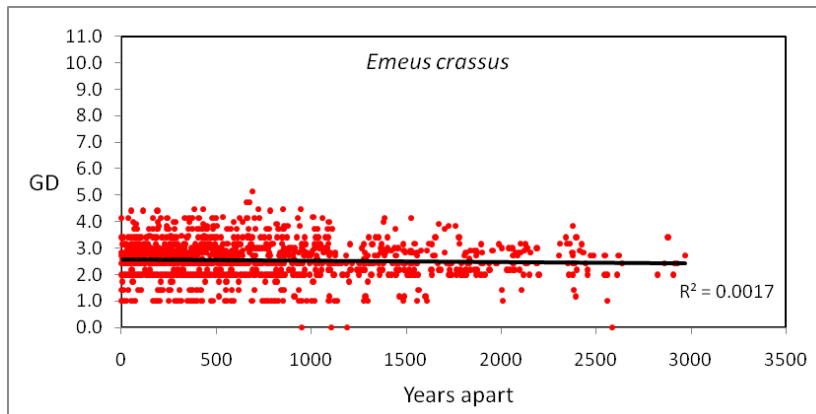


Figure 5.5d



5.2.7 Genetic relatedness and kinship among *Dinornis robustus*

Two methods were applied in an attempt to identify relatives in the data: a maximum likelihood based approach, as implemented in the software ML-RELATE (Kalinowski et al., 2006) and a more qualitative *ad hoc* assessment of relatedness-values in relation to radiocarbon dates.

ML-RELATE applies a maximum likelihood method based on relatedness-values to estimate if each pair of individuals have a higher probability of being unrelated, full siblings, half siblings or parent-offspring. Representing the largest and most informative dataset, a run was conducted on the *Dinornis robustus* data. All possible pairwise relationships were assessed according to a 95% significance level and 10000 permutations. A surprisingly high number of relatives were detected in the sample. Of 2701 pairwise relationships tested among the 74 individuals, ML-Relate estimated 551 family relations (71 full sib, 344 half sib and 136 parent-offspring relationships). The number was unrealistically high, given the >4000 year timeframe covered by the *Dinornis* data. To assess the validity, a simple ^{14}C age comparison was conducted based on the argument that truly related individuals must on average show a closer temporal connection than unrelated ones. The average age separation between suggested full siblings was 988 years, whereas the same numbers were 1019 years and 1021 years for parent-offspring and unrelated individuals respectively. Although the numbers indicated a weak tendency towards higher age-separation in unrelated individuals, it was not significant (t -test, $P = 0.50$). In short, it seems that ML-RELATE seriously overestimated the number of family relations in the data, and the method was not attempted on the less informative data representing the other species.

Next, potential family relations in the *Dinornis* data were examined with a more simplistic qualitative approach. Relatedness (R_{xy}) measures the relative genetic identity (identity by descent) among

individuals in a sample. Compared to the absolute genetic distances used in the IBT analysis above, the relatedness provides a more ancestrally related measure of genetic distance, and can be used to identify closely related individuals within a population, if the data are informative (e.g., Blouin, 2003). Pairwise R_{xy} values were calculated between all pairs of *Dinornis robustus* individuals according to Queller and Goodnight (1989) in GenAlEx ver.6 (Peakall, Smouse, 2006). No correlation was found when plotted against temporal distance (Figure 5.6), confirming the lack of signal in the IBT analysis above (Figure 5.5).

More interestingly, on four occasions did a high pairwise R_{xy} represent two individuals of ‘similar’ radiocarbon age (Figure 5.6). Additionally, in three of the four pairs, did the two individuals also share the same (relatively rare) mtDNA haplotype (Table 5.6, Figure 5.7). With three independent confirmations of tight genetic and temporal links, it seems highly likely that true relatives have been identified. The two individuals representing the fourth pair, did not share the same mtDNA haplotype (Table 5.6), eliminating a couple of possible family relations (full siblings, mother-offspring). All eight implicated individuals were adult females excavated from Pyramid Valley (Table 5.6). This finding of related adult females excavated from the same site, may represent important information on moa biology (see later discussion).

Figure 5.6: Relatedness in *Dinornis robustus*

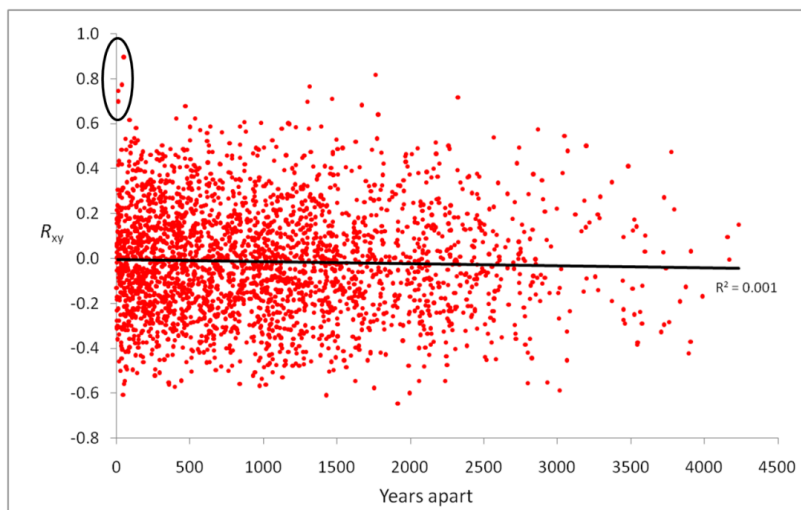


Figure 5.6: Values of relatedness (R_{xy}) between all pairs of individuals in the *Dinornis* data, plotted against the pairwise distance in time. No IBT signal could be observed with this measure either, but the circle identifies four pairs, in which the two individuals show high R_{xy} values and are of similar age.

Table 5.6: List of the four pairs of *Dinornis* individuals with high relatedness (R_{xy}) and close temporal associations, identified in Figure 5.6. In three of the four pairs, the two individuals also shared the same mtDNA haplotype. Haplotype IDs (A, B, C) refer to Figure 5.7, and the fraction of individuals sharing the same haplotype in the total *Dinornis* sample is indicated for each pair. Median calibrated radiocarbon age is provided for each individual with 95% confidence interval (see Chapter 2, Section 2 for explanation). PV = Pyramid Valley, F = female, “Maturity” refers to the ontogenetic classes defined in Turvey and Holdaway (2005). In Chapter 2 (Section 1), maturity classes V and VI were considered as adults. Although two of the individuals could not be genetically sexed, the sampled tibiotarsi were large (905 and 815 mm) showing that they were from females.

Museum #	R_{xy}	Haplotype	Age	Age, 95%	Site	Sex	Maturity
CM AV 8471	0.896	A: 4/74	2406 BP	2486-2342 BP	PV	F	VI
CM AV 8493			2357 BP	2460-2340 BP	PV	—	VI
CM AV 8466	0.773	not related	1766 BP	1825-1700 BP	PV	—	VI
CM AV 8486		through mother	1801 BP	1868-1730 BP	PV	F	VI
CM AV 8468	0.745	B: 5/74	2331 BP	2353-2183 BP	PV	F	VI
CM AV 8547			2318 BP	2353-2160 BP	PV	F	VI
AMNH 7301	0.699	C: 2/74	1326 BP	1380-1289 BP	PV	F	V-VI
CM AV 8475			1313 BP	1349-1288 BP	PV	F	VI

Figure 5.7: mtDNA median joining network for *Dinornis robustus*

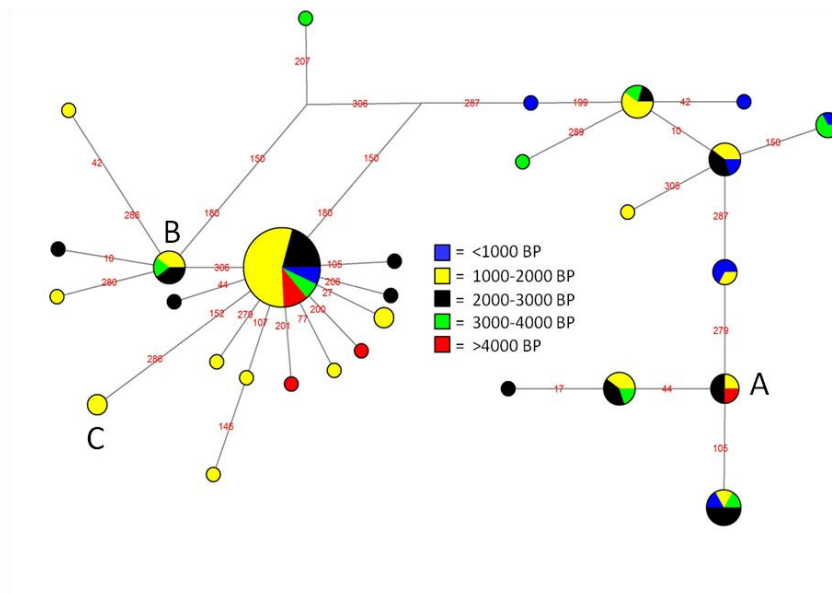


Figure 5.7: The mtDNA haplotype network for *Dinornis robustus* as presented in Chapter 4. A, B and C indicate the three haplotypes shared by the related individuals (see Table 5.6). Note the relative rarity of these three haplotypes, emphasising the improbability that the same individuals showing very tight temporal and multi-allelic associations would also cluster together by chance within these mtDNA haplotypes.

5.2.8 Identifying genetic outliers/migrants

Microsatellite data offer high resolution power to detect genetic outliers and the potential to identify immigrants and measure dispersal (Berry et al., 2004; Manel et al., 2005). To examine the data for genetically distinct individuals, a self-assignment test was conducted in GenAlEx ver. 6 (Peakall,

Smouse, 2006), providing each individual with a log-likelihood, measuring the strength of its genetic affiliation (Paetkau et al., 1995; Paetkau et al., 2004).

It was not possible to identify any finite lines between genetic outliers and non-outliers within the whole continuum of estimated likelihood values. Table 5.7 simply presents the five most genetically distinct individuals from each species. Some of these could potentially have been migrants from different subpopulations. *Dinornis* AV8470, for example, proved genetically least affiliated in this self-assignment test (Table 5.7) and this is the same individual, recognised as an outlier in the STRUCTURE analysis above (Figure 5.3). However, as with the identification of relatives, an independent confirmation from another source is required to verify the microsatellite signals. It was therefore investigated whether these 20 individuals also displayed unusual mtDNA profiles. The results shown on Figures 5.8a-d do not provide clear answers. Five of the 20 individuals (20 %) listed in Table 5.7 exhibited a unique mtDNA haplotype, which is only slightly higher than the 16 % expected by chance*. In other words, there are no convincing signs that the two datasets showed congruence in terms of potential migrants.

*Expected fraction is simply found as:

- Fraction of individuals with a unique mtDNA haplotype in the *Dinornis* data: 18/87 individuals = 20.7%
- Per five individuals (as in Table 5.7) this was: $5 \times 0.207 = 1.035$ individuals
- The same number was calculated for the three other species and when added up, it yielded 3.26 individuals, corresponding to 16 % of 20 individuals

Table 5.7: The five individuals (listed by museum catalogue numbers) from each species that exhibited least genetic affiliation with their respective gene pools. Log-likelihood values are shown, with more negative values representing weaker affiliation. The log-likelihood mean for each species is shown in brackets next to taxon names.

#	<i>D. robustus</i> (-7.02)	<i>P. eleph.</i> (-6.52)	<i>E. curtus</i> (-6.64)	<i>E. crassus</i> (-2.89)
1	AV 8470 -11.59	AV 8388 -9.79	S 40257 -9.97	AMNH 7306 -4.79
2	AV 8491 -11.03	AV 15029 -9.70	S 40157 -9.19	S 39934-6 -4.75
3	AV 15028 -10.75	S 40344 -9.69	AMNH 7304 -9.00	AV 13774 -4.65
4	S 40330 -9.62	AV 8385 -8.79	S 15035 -8.86	AV 8306 -4.62
5	AV 8477 -9.61	S 40021 -8.12	S 39934-12 -7.70	S 40386 -4.58

Figure 5.8a: *Dinornis robustus*

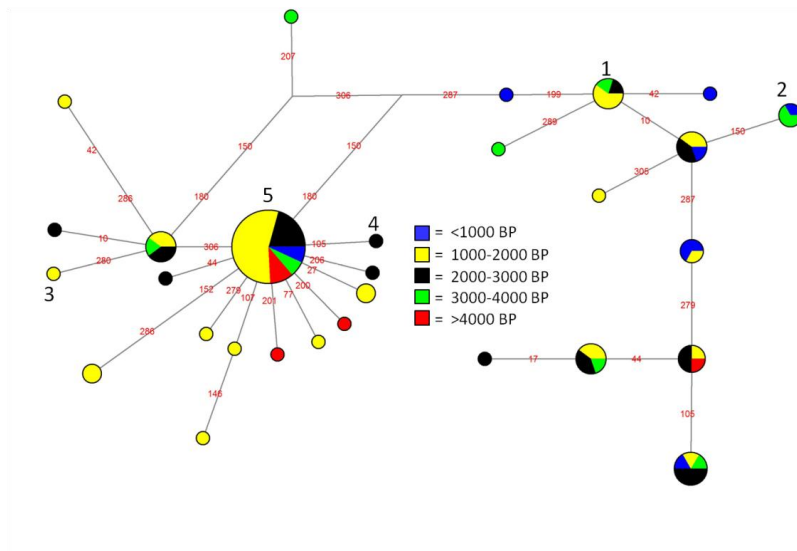


Figure 5.8a-d: Median-joining networks based on mtDNA, as presented in Chapter 4. The position of individuals with unusual microsatellite profiles are indicated with numbers corresponding to Table 5.7.

Figure 5.8b: *Pachyornis elephantopus*

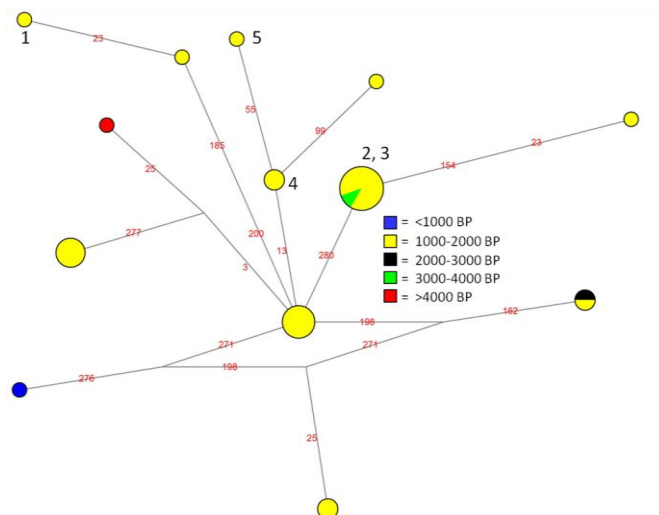


Figure 5.8c: *Euryapteryx curtus*

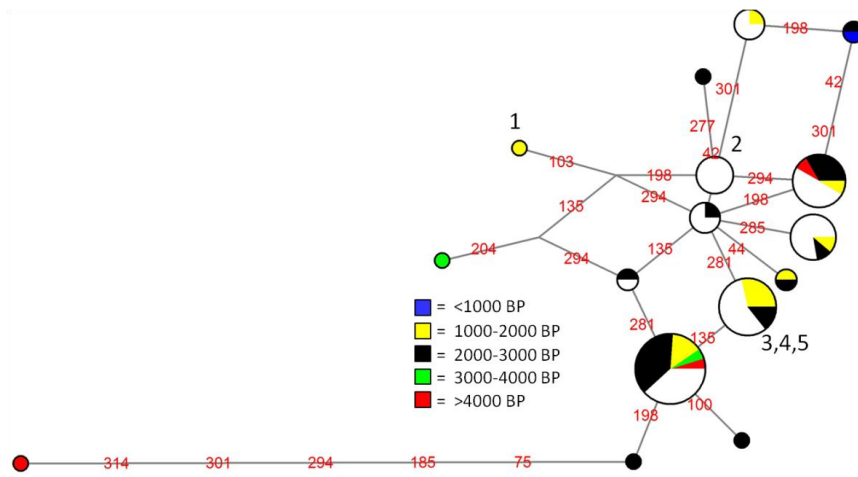
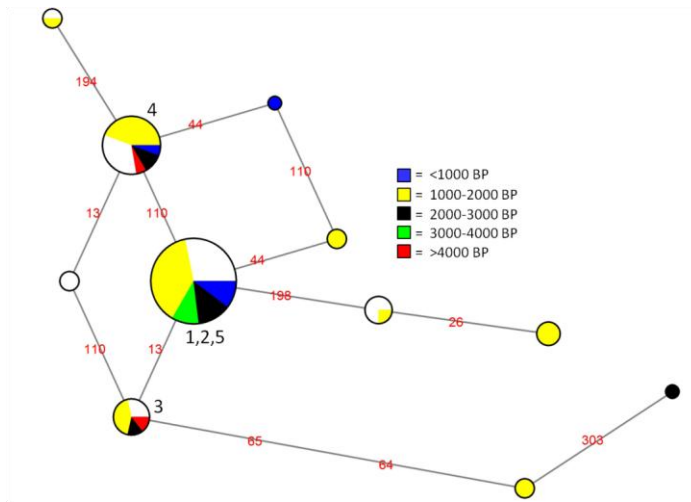


Figure 5.8d: *Emeus crassus*



5.2.9 Sex biased dispersal

Three genetic data-components constitute the work presented in this thesis (sex ID, mtDNA and microsatellites). When investigated in concert, they represent a potential for exposing patterns in inter- and intraspecific genetic diversity at many different levels. To investigate the hypothesis of male biased dispersal, measures of genetic diversity and relatedness were calculated for males and females respectively, using both mtDNA and microsatellite data (Tables 5.8 and 5.9). If males would tend to disperse further than females, then males should, in theory, display lower relatedness and higher genetic differentiation than the local females. Juveniles, regardless of gender, may not have dispersed at the time of death, so they were here analysed in a separate group to prevent obscuring any potential male-female differences. Measures of mitochondrial haplotype diversity were calculated in DnaSP (Librado, Rozas, 2009), whereas microsatellite diversity and average relatedness values were calculated in GenAlEx ver. 6 (Peakall, Smouse, 2006) according to Queller and Goodnight (1989). The robustness of the reported within-group mean relatedness (R_{xy}) was assessed with 1000 bootstraps and tests for significance of R_{xy} (higher than among individuals with randomly shuffled alleles) were conducted with 1000 permutations.

No obvious differences emerged between the groups (Tables 5.8 and 5.9). Adult males of *Euryapteryx curtus* may display lower mtDNA diversity than the females (Table 5.8), but this could not be confirmed by microsatellite data, which only included only one adult male of this species. There were no differences in relatedness among groups (Figure 5.9a-d) except for six males of *Pachyornis* showing higher average relatedness ($R_{xy} = 0.179$) than the females ($R_{xy} = -0.053$) (Table 5.9, Figure 5.9b), and the relatedness among these six males was bordering significance ($P = 0.06$) in the

permutation test (i.e. higher than expected by random). If correct, this would refute the idea of males dispersing wider than females, but the patterns are very inconclusive. There is no secure evidence that males, females and juveniles, sampled from this small geographic area, reflected the gene pool differently.

Table 5.8: Measures of mtDNA (Control region) genetic diversity as presented in Chapter 4, but here separated on adult males, adult females and juveniles. Sample sizes (n) are defined by number of individuals that have a Control region sequence, a molecular gender identification, and a maturity stage identification (as outlined in Chapter 2). H_d is haplotype diversity and π is nucleotide diversity.

	adult females			adult males			juveniles		
	n	H_d	π	n	H_d	π	n	H_d	π
<i>D. robustus</i>	44	0.862	0.008	12	0.848	0.008	13	0.905	0.011
<i>P. elephantopus</i>	25	0.883	0.008	3	1.000	0.005	2	1.000	0.008
<i>E. curtus</i>	26	0.917	0.008	6	0.533	0.003	40	0.858	0.007
<i>E. crassus</i>	55	0.706	0.004	9	0.833	0.003	10	0.8	0.004

Table 5.9: Measures of microsatellite genetic diversity (H_E = expected heterozygosity) and mean relatedness (R_{xy}), separated on adult males, adult females and juveniles. Sample sizes (n) are defined by number of individuals with both a complete multi-locus genotype, a successful molecular gender identification, and a maturity stage identification (as outlined in Chapter 2).

	adult females			adult males			juveniles		
	n	H_E	R_{xy}	n	H_E	R_{xy}	n	H_E	R_{xy}
<i>D. robustus</i>	43	0.718	-0.007	8	0.745	-0.046	17	0.723	-0.011
<i>P. elephantopus</i>	23	0.609	-0.053	4	0.518	0.179	2	0.639	-0.367
<i>E. curtus</i>	12	0.579	-0.051	1	NA	NA	16	0.555	-0.006
<i>E. crassus</i>	39	0.290	-0.004	6	0.275	0.075	8	0.287	-0.022

Although relatedness and genetic diversity separated on genders did not reveal any specific patterns, there are a range of methods available to investigate gender related dispersal patterns (reviewed in Prugnolle, de Meeus, 2002). These include the sex bias assignment test (Favre et al., 1997; Mossman, Waser, 1999) implemented in GenAlEx ver. 6 (Peakall, Smouse, 2006). The test is based on allele frequencies and provides each individual with a relative assignment index correction (Aic) value, reflecting the likelihood of being an immigrant (the more negative value, the more likely to be an immigrant). Gender biased dispersal is expected to generate different Aic values in males and females. For example, if male moa were predominantly from different subpopulations (with different allele frequencies) they should generally display lower Aic values than females.

Figure 5.9a

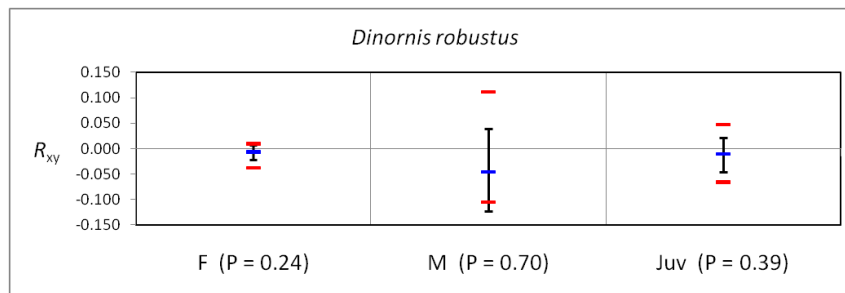


Figure 5.9b

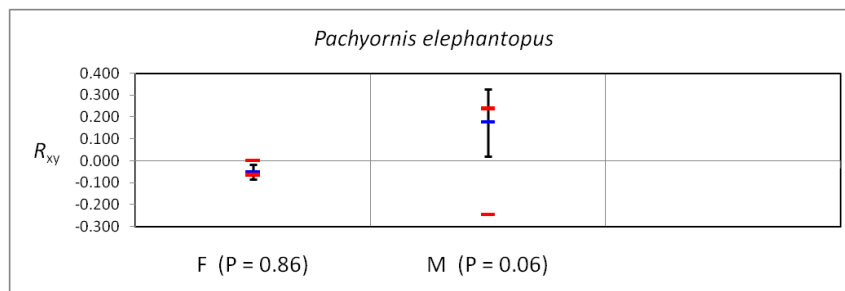


Figure 5.9c

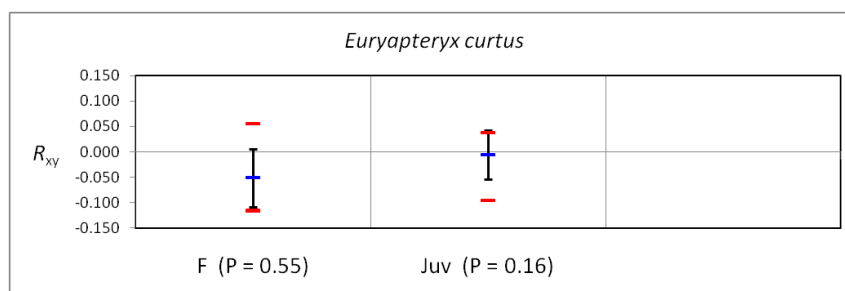


Figure 5.9d

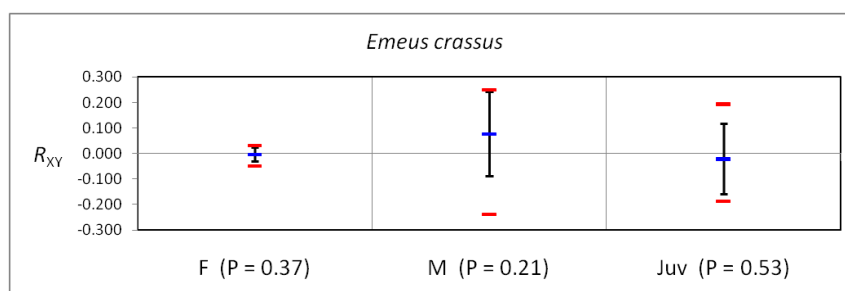


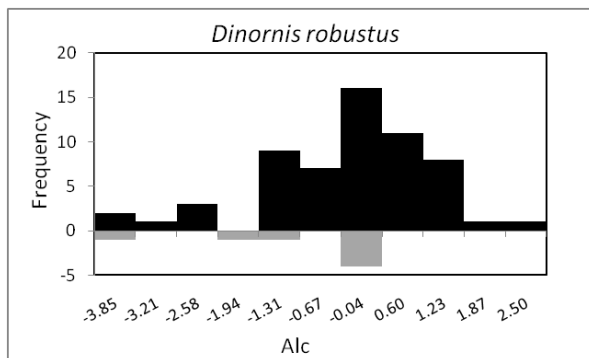
Figure 5.9a-d: Pairwise genetic relatedness within males, females and juveniles from each species respectively. Blue markers represent mean values as listed in Table 5.8. Error bars bound the 95% confidence intervals around the mean values (determined by bootstrap re-sampling). Red markers bound the 95% confidence interval of a null hypothesis of 'no difference' across the populations (determined by permutation). *P*-values represent the probability of the observed mean being larger than the permuted mean within each group.

Results from *Dinornis*, *Pachyornis* and *Emeus* are shown in Figures 5.10a-c. Only one adult *Euryapteryx* male had been genotyped, making the analysis inconclusive for this species. The Aic-

frequency distributions were largely overlapping between genders, perhaps indicating that both sexes were dispersing equally.

There may be a tendency for male *Pachyornis* to show stronger North Canterbury affiliation than an average female from this taxon, corresponding to the higher male relatedness shown above. But small sample sizes and heterochronous data may hamper these interpretations.

Figure 5.10a



Figures 5.10a-c: Investigating sex biased dispersal in moa. Aic-frequencies are largely overlapping between sexes, indicating that dispersal is not predominantly facilitated by one sex.

Black columns = females

Grey columns = males

Figure 5.10b

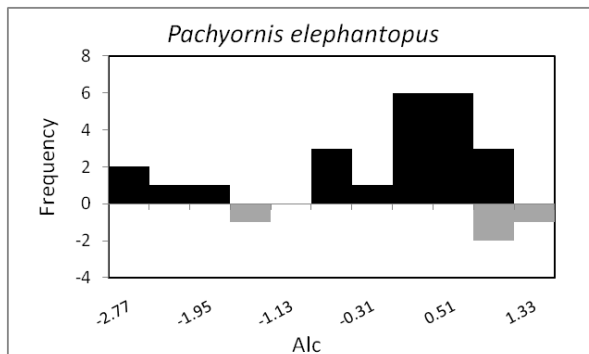
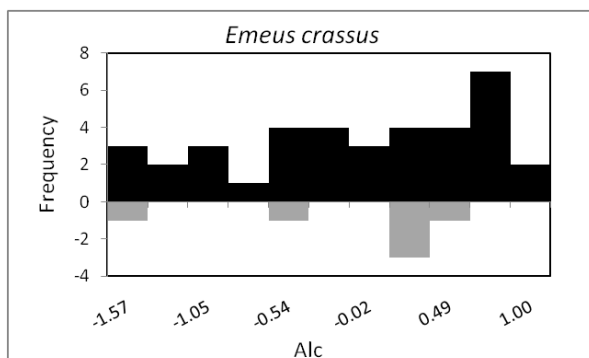


Figure 5.10c



5.3 Discussion

This chapter describes genetic analyses based on six polymorphic microsatellites markers and 188 genotyped individuals from four extinct taxa, covering 4000 years in the Holocene. The data presented here are unique in aDNA research. The success rate has once again confirmed the extraordinary biomolecule preservation in these bones. Although several of the analyses are equivalent to basic exploratory investigations of microsatellite data, as encountered thousands of times in literature on extant species, the framework is without precedence. Therefore, in addition to a primary objective of gaining new detailed information on moa, the secondary objective was to explore the potentials and reveal the strengths and the caveats with this type of data. In this context, it was obvious that the work involved some significant analytical obstacles, here summarised in three points:

- 1) Although having data from six polymorphic loci, in four extinct species, is beyond anything previously achieved or even attempted in the field of aDNA, it is still a relatively low number of loci in terms of resolution power. In the end, the resolution depends heavily on the genetic variability in the markers, but in most present-day microsatellite population studies, 10-20 markers are used.
- 2) Temporal genetic variation in this data could potentially mask any signals, by creating too much background noise for the analyses to be effective. In microsatellite analyses, levels of significance and likelihood values etc. are commonly calculated based on allele frequencies, presumed to reflect the population at one particular point in time. With heterochronous data, the reference point is not a 'point', but essentially an assemblage of genetic snapshots, collected from many different points in time.
- 3) Lastly, the fact that all samples have been excavated from the same area is in many regards a significant advantage, but at the same time, the true power of microsatellite data often comes to expression when revealing spatial patterns in genetic variability. With single-population data, the possibilities are somewhat more limited.

Listing the three points here is not intended to discredit the analyses or otherwise compromise the results presented in this chapter. However, since this dataset is the first of its kind, it is important to realise the careful considerations and reservations underlying the interpretations. To briefly address the three issues; the first point is a matter of resources as already discussed in Chapter 3 (Section 2). It is not possible to generalise, but the dataset for *Dinornis robustus*, in particular, was large and displayed a high level of genetic variability. Such data would probably qualify as being very informative in most analytical context. Point 2 and 3 are more interesting in that the problematic effects mentioned here, may cancel each other out. In this project, genetic diversity has effectively only been monitored in one dimension, namely time. Therefore, the problems of discriminating between temporal and spatial variability may be more in the mind of the geneticist, trained through

conventional “spatial genetics” than actual limitations offered by the data. This topic is revisited several times below, where the results of the microsatellite analyses will be discussed according to the six points from the introduction.

5.3.1 Basic descriptive statistics, tests for HW-proportions and LD

Apart from one taxon-locus combination, the data displayed no notable trends in terms of deviations from HW-proportions and LD. This result was unexpected, given that individuals across 4000 years had been pooled. The data suggested that the allele frequencies remained more or less constant throughout the entire last half of the Holocene. Thus, the HW- and LD tests provided the first indications of cohesive gene pools, showing that the data could be analysed together, rather than split up into temporal (or spatial) bins.

In terms of genetic diversity, the microsatellite data revealed patterns, similar to the mtDNA in Chapter 4. *Dinornis robustus* showed highest genetic diversity, although there may be some ascertainment bias affecting the result. Regardless, there was convincing evidence that the *Emeus crassus* population displayed much lower genetic diversity than the other three taxa. This is a confirmation of the hypothesis outlined in Chapter 4 and Bunce et al. (2009), suggesting that this species had experienced a major bottleneck before the mid-Holocene, for example as a result of the last glacial period. This congruence between the mtDNA and microsatellite data was reassuring and confirmed the overall fidelity of the two datasets.

The concept that populations with low levels of genetic variability have higher extinction risk is well established (Frankel, Soulé, 1981; Frankham et al., 2002; Reed, Frankham, 2003; Shaffer, 1990; Spielman et al., 2004). Individuals in populations with low genetic variation have a higher probability of becoming genetically inbred (autozygous for deleterious alleles), with the potential consequence of lowered fitness. Additionally, the inherent genetic variability of populations is considered adaptive to changing environmental conditions, thereby acting as a buffer against stochastic and catastrophic events (e.g., Frankham et al., 2002; Willi et al., 2006). In that light, it is interesting that *Emeus crassus* was not only able to persist, but even appeared to be thriving on the Eastern Plains, where Holocene fossils have been found in great abundance (Holdaway, Worthy, 1997; Worthy, Holdaway, 1996; Worthy, Holdaway, 2002). There is, of course, no universally accepted critical threshold level for low genetic diversity. Microsatellites are neutral loci so the variability can only be used as an indirect relative measure of the variability in genes that are under selection (DeWoody, DeWoody, 2005; Hansson, Westerberg, 2002). Moreover, heterozygosity is not always easily compared between species, unless the same loci have been amplified and analysed. Even in cross-species studies, using

the exact same loci, the problem of ascertainment-bias may compromise the interpretations (Goldstein, Pollock, 1997). Despite these caveats, it is clear that *Emeus crassus* maintained about half the heterozygosity of the other moa species for at least 3000 years (Table 5.2, Figure 5.2d). If taken out of context, this result could be interpreted as a serious challenge to the entire field of “Conservation Genetics”. However, in terms of inbreeding depression, it is essential to discriminate between rapid genetic bottlenecks, for example mediated by recent anthropogenic landscape modifications (often the target of conservation surveys), and low genetic diversity echoing ancient demographic or phylogeographic events. In the latter case, natural selection may have had of time to “purge” mildly deleterious alleles, meaning that low genetic diversity is not necessarily accompanied by lowered fitness in the population (Frankham et al., 2002; Hedrick, 1994; Leberg, Firmin, 2008). It is not unlikely that purging has been responsible for the apparent robustness of *Emeus* despite its low genetic diversity. The idea of a long standing small population, isolated in a glacial refugium for thousands of years, followed by a post glacial expansion has already been invoked in Chapter 4. Aside from the direct fitness related effects (i.e. inbreeding depression), low genetic diversity is also expected to lower the adaptive potential. So the fact that *Emeus* was able to persist for at least 3000 years with relatively low genetic diversity, is perhaps an indication of a stable environment without sudden shifts in selective pressure. This hypothesis of a stable environment in North Canterbury throughout the late Holocene is strongly supported from stable isotopic data (Holdaway et al. in prep).

In population genetic studies, it is common practice to compare the documented level of genetic diversity with that of other related species. Despite the potential caveats in these comparisons, they can provide an appropriate context for a further discussion on the genetic signals, especially in closely related species, such as the three emeids (*Pachyornis*, *Euryapteryx* and *Emeus*). The best possible interspecific comparison has already been carried out, by considering four moa species together here. However, if extending the focus to include ratites in general, the lack of comparable studies is quite astounding – in particular when considering the availability of numerous ratite microsatellites markers (Brennan, Hyseni, 2008; Kimwele et al., 1998; Roots, Baker, 2002; Tang et al., 2003; Taylor et al., 1999; Ward et al., 1998). In fact, ratite population genetics can be summarised as: one study on the genetic structure of greater rhea, from a small area in Argentina (Bouzat, 2001) based on RAPD markers; one study of mtDNA diversity in ostrich based on restriction enzyme chemistry (Freitag, Robinson, 1993); one assessment of microsatellite diversity within, an among, ostrich breeds used in Polish farming (Kawka et al., 2007); and two microsatellite studies on nesting behaviour and parentage analyses in ostrich and emu (Kimwele, Graves, 2003; Taylor et al., 2000). The New Zealand kiwi (*Apteryx* spp.) has been studied in somewhat greater detail (Baker et al., 1995; Burbidge et al., 2003; Herbert, Daugherty, 2002; Jolly, Daugherty, 2002; McLennan, McCann, 2002; Shepherd, Lambert, 2008) but none of these studies used microsatellites. Of the listed publications, only Kimwele and Graves (2003) provided data, comparable with this moa study. Based on the genetic

variation in eight microsatellites, the authors analysed the complex communal nesting strategy in ostriches. In terms of genetic diversity, H_O varied from 0.61 to 0.84 between markers. Although not listing an overall mean value, it was slightly higher than documented here in moa (Table 5.2). Within moa, H_O was highest in *Dinornis robustus*, varying between 0.60 and 0.79, with a mean across loci of 0.69 (Table 3.2.4).

With reference to Table 1.1.1, it is somewhat ironic that the genetics of a long gone extinct lineage has been studied with much greater effort than all the extant ratites combined. This, however, makes it difficult to assess the relative levels of genetic diversity beyond the intra-specific comparison that has already been discussed. Further research in ratite population genetics, especially kiwi (*Apteryx* spp.), which shared a similar habitat, is required and may shed more light on the data presented here.

5.3.2 Genetic diversity through time

Considering the large number of softwares packages that are available for generating simulated microsatellite data, for example by assuming random mating through x-number of generations (reviewed in Carvajal-Rodriguez, 2010), and considering how many theoretical predictions and publications that have been based on simulated microsatellite data, it is interesting to observe such data from real populations. By applying a predicted generation time of 10 years (as outlined in Chapter 4), this moa dataset is likely to cover ~400 generations. Unfortunately, this is also a good example of unique data that are very difficult to interpret. As seen on Figures 5.2a-d, there is large individual variation in heterozygosity at any point in time, which can, in part, be ascribed to the relatively few markers used. This implies that there is a large sampling error involved, making it difficult to estimate a population mean-value through time. Yet, there are no indications in these data for a loss of heterozygosity through time, as would be expected if indeed a population in decline had been sampled, as suggested in Gemmell et al. (2004).

The correlation between population size and genetic diversity is not only a theoretical prediction but has been shown many times empirically, for example in a large influential meta-analysis by Frankham (1996). In terms of documenting a loss of genetic diversity in microsatellite data, numerous studies have used contemporary samples to present signals of past genetic bottlenecks, based on retrospective models and theoretical predictions (e.g., Garza, Williamson, 2001; Piry et al., 1999). However, far fewer studies have utilised temporal sampling, which is required to directly document and measure such changes in genetic variability over time. By sampling museum specimens of Darwin's finches, Petren et al. (2010) analysed the genetic information in 14 microsatellite loci to document a considerable loss of genetic diversity from the time of the Beagle Voyage and to present day.

Whitehouse and Harley (2001) analysed the genetic variation in nine microsatellite loci from >200 South African elephants and included two museum specimens from 1920. The old specimens displayed several alleles not present in the modern samples. Similarly, evidence for loss in microsatellite diversity has been documented in Mauritius kestrel (Groombridge et al., 2000), greater Prairie-chicken (Johnson et al., 2004), and Adonis blue butterfly (Harper et al., 2006). Although none of these studies covered a temporal window comparable to the moa data, they showed that temporal microsatellite data offer a powerful tool to detect loss of genetic diversity or temporal changes in allele frequencies. Taken together, the literature suggests that any major bottlenecks in the moa populations during the sampled time frame would be unlikely to have gone completely undetected, with the data and analyses in this chapter.

5.3.3 Genetic structure and isolation by time

It is well documented that STRUCTURE-results should be interpreted with caution, for example because the assumption of random mating is often violated in natural populations. The authors urge STRUCTURE-users to apply biologically meaningful interpretations rather than trusting blindly the K -value showing highest likelihood (Falush et al., 2003; Pritchard et al., 2000). Simulation studies have been carried out and methods have been developed to assist in identifying the ‘true’ K -value (e.g., Evanno et al., 2005). In the case of the moa, the results were so clearly displaying a lack of genetic structure that any further elaboration seemed unnecessary. The data displayed a remarkable homogeneity across the entire time period. Even when violating the highly variable *Dinornis robustus* data, by excluding individuals to generate a 2000 year gap in the sample, the result still showed complete admixture, as if individuals have been drawn from exactly the same gene pool. This genetic homogeneity was only further supported by the low F_{ST} values and a lack of an IBT signal. Remarkably, the genetic distance between individuals thousands of years apart was no larger than between contemporary individuals.

Together, these analyses provided clear evidence that the effect of genetic drift was extremely limited over the 4000 years monitored here. The obvious interpretation of such signals of ‘genetic stasis’ would be that they reflect very large populations. Although simulation-based analyses (discussed further below) are probably required to definitively reject the population decline hypothesised in Gemmell et al. (2004), it seems highly unlikely that *Dinornis robustus* could have experienced a major decline in population size during the late Holocene, without displaying loss of genetic diversity (in mtDNA or microsatellite data) or even subtle changes in allele frequencies due to drift.

A comparison of the genetic structure in the two genetic datasets (microsatellites and mtDNA) representing *Dinornis*, proved highly informative. The microsatellite data showed a completely homogenous gene pool, whereas the mtDNA revealed two distinct haplogroups (and a single distinct individual) (Figure 4.2a). This is an example of the advantages in applying several types of molecular markers in concert. From the mitochondrial DNA, it seems likely that several old maternal lineages were present in the North Canterbury *Dinornis* population, possibly because the area represented a secondary contact zone in the Holocene, reflecting postglacial dispersal patterns (discussed in Chapter 4). The microsatellites on the other hand, provided a much more ‘contemporary’ image of the population, and revealed a very homogenous gene pool through the last 4000 years of the moa era.

As a final ‘twist’, the STRUCTURE analyses demonstrated that the microsatellite markers can work as species identification tools. This application is more of academic than practical importance, considering the work effort compared to the established mtDNA protocols. However, the clearly defined STRUCTURE result at $K=4$ (Figure 5.3) served as a solid independent confirmation of the species identifications provided in Chapter 2.

5.3.4 Kinship analyses

It took a qualitative *ad hoc* method and the integration of three different datasets to securely identify relatives among the *Dinornis* fossils. It may indeed have biological significance that three or four pairs of adult related females were deposited at Pyramid Valley. Without engaging in long speculation on a limited background, the result could suggest that related individuals of *Dinornis* (adult females at least) stayed together in flocks or pairs, or perhaps more realistically that they did not disperse far from the area where they hatched. The alternative hypothesis would be that moa dispersed widely, but that adult related individuals would occasionally, by sheer chance, die at the same site. This is a fully plausible scenario, but the chance of that happening four times at Pyramid Valley, given the deposition rates, seems small. There were indeed many other places for a moa to die in Canterbury and there were many moa individuals in the population to “pick from” (around 10^5 according to the estimate in Chapter 4). Given these kinship results, combined with the hypotheses of female territoriality presented in Chapter 2 (Section 1), the favoured interpretation is that females matured and established a home range close to the area where they hatched. This does not necessarily imply that males did not do the same.

It is not possible with this method to determine how the individuals were related. As mentioned, the two individuals constituting one of the four pairs could not have had the same mother. If assuming that the high R_{xy} value and lack of temporal separation is secure evidence for kinship, then this is an

indication that males would mate with different females throughout their life. Extant ratites are known to have complex mating structures but varying extensively among the different species (e.g., Kimwele, Graves, 2003; Taylor et al., 2000), which makes it difficult to use as reference for moa behaviour. However, within ratites, monogamous behaviour is known only from kiwi (Taborsky, Taborsky, 1999) and the results presented in this chapter could suggest that moa were not monogamous either.

5.3.5 Sex biased dispersal and genetic outliers

These analyses did not result in any unequivocal results. The accuracy of assignment tests is reduced when only one population has been sampled and when limited genetic variability prevails (Manel et al., 2005; Piry et al., 2004; Prugnolle, de Meeus, 2002). However, rather than being a problem with the data, the result may very well reflect a large homogenous gene pool, with only the really rare long distance migrants standing out in an assignment test. Two outliers were detected in the mtDNA dataset: one distinct *Dinornis* (Haplogroup C on Figure 4.2a) and a single very distinct *Euryapteryx* from Rosslea (Figure 4.2d). The *Dinornis* individual did not appear among the least affiliated microsatellite genotypes, and the DNA preservation in the *Euryapteryx* from Rosslea was too poor for microsatellite work.

5.4 Concluding remarks

This study represents the first high resolution population genetic survey of extinct species. The two major findings can be summarised as i) large differences in genetic diversity between the four sympatric populations and ii) an apparent genetic stasis of all four gene pools throughout the 4000 years leading up to moa extinction. Had this been spatial data, sampled at one point in time, the conclusion would clearly be that all individuals had been sampled from the same population: panmixia. It is interesting to transfer this interpretation to a temporal dimension. Clearly, there can be no such thing as ‘temporal panmixia’ across a 4000 year timeframe, but it is essentially what the data shows. It seems that genetic drift was so limited that the present alleles had an equal chance of being associated across the entire temporal window examined. With no effects of genetic drift observed, it would suggest that the moa populations were very large, or were maybe even increasing in size, as indicated for *Dinornis robustus* in the Chapter 4 (Figure 4.3a). All the results presented in Chapters 4 and 5 have suggested that the four North Canterbury moa were part of large viable populations throughout the late Holocene. Around 600 BP (1350 CE) deposition stopped, and the extinction was so rapid that it did not manifest itself in the moa gene pools.

However, despite a series of strong indications favouring the opposite, it is not possible here to definitively reject the pre-human Holocene decline scenario from Gemmell et al. (2004). This is mainly because the population sizes and demographic histories have not yet been assessed with these microsatellite data. A range of methods are available for estimating population sizes (reviewed in Luikart et al., 2010; Schwartz et al., 1998) and detecting bottlenecks (Cornuet, Luikart, 1996; Garza, Williamson, 2001; Piry et al., 1999) based on allele frequencies, but none of these apply to data that cover thousands of years. Instead they rely on accurate estimates of population allele frequencies, sampled at one or two points in time. Although the applicability remains to be fully investigated, it seems that using these ‘simple’ N_e estimators to the moa microsatellite data, would severely violate the underlying assumptions, equivalent to the previous population size estimate based on mtDNA (Gemmell et al., 2004). Given these reservations, an analysis of moa population sizes using microsatellite data has not yet been carried out.

This point brings the discussion to an end with the two major questions that still remains to be answered:

- 1) What level of resolving power do these microsatellite data offer? Would they display signals of demographic fluctuations if they had indeed occurred, for example by recording a lowered H_e or changes in allele frequencies through time?
- 2) Could the moa populations have maintained the recorded level of genetic diversity through a severe decline? In other words; would the data in Figure 5.2 look similar if moa had experienced the scenario outlined in Gemmell et al. (2004)?

Theoretical data-simulation studies are likely to be the way forward and several methods are available for generating artificial data under given demographic scenarios (reviewed in Carvajal-Rodriguez, 2010). A couple of methods based on approximate Bayesian computation (ABC) seem particularly promising (Anderson et al., 2005; Cornuet et al., 2008), and will be thoroughly explored in the near future. In addition, a coalescent based method, building genealogies from microsatellite data, is under development (Alexei Drummond, pers. comm.), which means that in time, it may be possible to analyse this data in a Bayesian framework, using Bayesian Skyline plots, similar to that seen with mtDNA.

5.5 References

- Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA (2005) Serial SimCoal: A population genetics model for data from multiple populations and points in time. *Bioinformatics* **21**, 1733-1734.
- Baker AJ, Daugherty CH, Colbourne R, McLennan JL (1995) Flightless brown kiwis of New Zealand possess extremely subdivided population structure and cryptic species like small mammals. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 8254-8258.
- Berry O, Tocher MD, Sarre SD (2004) Can assignment tests measure dispersal? *Molecular Ecology* **13**, 551-561.
- Birchard GF, Deeming DC (2009) Avian eggshell thickness: scaling and maximum body mass in birds. *Journal of Zoology* **279**, 95-101.
- Blouin MS (2003) DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology & Evolution* **18**, 503-511.
- Bouzat JL (2001) The population genetic structure of the Greater Rhea (*Rhea americana*) in an agricultural landscape. *Biological Conservation* **99**, 277-284.
- Brennan PLR, Hyseni C (2008) Development of microsatellite markers for parentage analysis in the great tinamou (*Tinamus major*). *Molecular Ecology Resources* **8**, 933-935.
- Bunce M, Worthy TH, Ford T, et al. (2003) Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425**, 172-175.
- Bunce M, Worthy TH, Phillips MJ, et al. (2009) The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20646-20651.
- Burbidge ML, Colbourne RM, Robertson HA, Baker AJ (2003) Molecular and other biological evidence supports the recognition of at least three species of brown kiwi. *Conservation Genetics* **4**, 167-177.
- Buschiazzi E, Gemmell NJ (2006) The rise, fall and renaissance of microsatellites in eukaryotic genomes. *BioEssays* **28**, 1040-1050.
- Carvajal-Rodriguez A (2010) Simulation of Genes and Genomes Forward in Time. *Current Genomics* **11**, 58-61.
- Clarke AL, Saether BE, Roskaft E (1997) Sex biases in avian dispersal: A reappraisal. *Oikos* **79**, 429-438.
- Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001-2014.
- Cornuet JM, Santos F, Beaumont MA, et al. (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**, 2713-2719.
- DeWoody YD, DeWoody JA (2005) On the estimation of genome-wide heterozygosity using molecular markers. *Journal of Heredity* **96**, 85-88.
- Dobson FS (1982) Competition for mates and predominant juvenile male dispersal in mammals. *Animal Behaviour* **30**, 1183-1192.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.

- Favre L, Balloux F, Goudet J, Perrin N (1997) Female-biased dispersal in the monogamous mammal *Crocivura russula*: Evidence from field data and microsatellite patterns. *Proceedings of the Royal Society of London Series B-Biological Sciences* **264**, 127-132.
- Frankel OH, Soule ME (1981) *Conservation and Evolution* Cambridge University Press, Cambridge.
- Frankham R (1996) Relationship of genetic variation to population size in wildlife. *Conservation Biology* **10**, 1500-1508.
- Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to Conservation Genetics* Cambridge University Press, Cambridge.
- Freitag S, Robinson TJ (1993) Phylogeographic patterns in mitochondrial DNA of the ostrich (*Struthio camelus*). *Auk* **110**, 614-622.
- Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* **10**, 305-318.
- Gemmell NJ, Schwartz MK, Robertson BC (2004) Moa were many. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, S430-S432.
- Goldstein DB, Pollock DD (1997) Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* **88**, 335-342.
- Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from: <http://www.unil.ch/izea/software/fstat>.
- Greenwood PJ (1980) mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour* **28**, 1140-1162.
- Groombridge JJ, Jones CG, Bruford MW, Nichols RA (2000) Conservation biology - 'Ghost' alleles of the Mauritius kestrel. *Nature* **403**, 616-616.
- Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology* **11**, 2467-2474.
- Harper GL, Maclean N, Goulson D (2006) Analysis of museum specimens suggests extreme genetic drift in the adonis blue butterfly (*Polyommatus bellargus*). *Biological Journal of the Linnean Society* **88**, 447-452.
- Hedrick PW (1994) Purging inbreeding depression and the probability of extinction - full sib mating. *Heredity* **73**, 363-372.
- Hedrick PW (2000) *Genetics of Populations* Jones and Bartlett Publishers, Inc., London.
- Herbert J, Daugherty CH (2002) Genetic variation, systematics and management of kiwi (*Apteryx* spp.). In: *Some early 1990s studies in kiwi (Apteryx spp.) genetics and management* (ed. Overmars F), pp. 11-33. Department of Conservation, Wellington, New Zealand.
- Holdaway RN, Worthy TH (1997) A reappraisal of the late quaternary fossil vertebrates of Pyramid Valley Swamp, North Canterbury, New Zealand. *New Zealand Journal of Zoology* **24**, 69-121.
- Huynen L, Millar CD, Scofield RP, Lambert DM (2003) Nuclear DNA sequences detect species limits in ancient moa. *Nature* **425**, 175-178.
- Johnson JA, Bellinger MR, Toepfer JE, Dunn P (2004) Temporal changes in allele frequencies and low effective population size in greater prairie-chickens. *Molecular Ecology* **13**, 2617-2630.
- Jolly JN, Daugherty CH (2002) Comparison of little spotted kiwi (*Apteryx owenii*) from Kapiti and D'Urville Islands. In: *Some early 1990s studies in kiwi (Apteryx spp.) genetics and management* (ed. Overmars F), pp. 57-64. Department of Conservation, Wellington, New Zealand.
- Kalinowski ST, Wagner AP, Taper ML (2006) ML-RELATE: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes* **6**, 576-579.
- Kawka M, Horbanczuk JO, Sacharczuk M, et al. (2007) Genetic characteristics of the ostrich population using molecular methods. *Poultry Science* **86**, 277-281.

- Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725-&.
- Kimwele CN, Graves JA (2003) A molecular genetic analysis of the communal nesting of the ostrich (*Struthio camelus*). *Molecular Ecology* **12**, 229-236.
- Kimwele CN, Graves JA, Burke T, Hanotte O (1998) Development of microsatellite markers for parentage typing of chicks in the ostrich *Struthio camelus*. *Molecular Ecology* **7**, 249-251.
- Leberg PL, Firmin BD (2008) Role of inbreeding depression and purging in captive breeding and restoration programmes. *Molecular Ecology* **17**, 334-343.
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.
- Luikart G, Ryman N, Tallmon DA, Schwartz MK, Allendorf FW (2010) Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conservation Genetics* **11**, 355-373.
- Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions techniques with appropriate. *Trends in Ecology & Evolution* **20**, 136-142.
- McLennan J, McCann T (2002) Genetic variability, distribution and abundance of great spotted kiwi (*Apteryx haastii*). In: *Some early 1990s study in kiwi (Apteryx spp.) genetics and management* (ed. Overmars F), pp. 35-55. Department of Conservation, Wellington, New Zealand.
- Mossman CA, Waser PM (1999) Genetic detection of sex-biased dispersal. *Molecular Ecology* **8**, 1063-1067.
- Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* **4**, 347-354.
- Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* **13**, 55-65.
- Park SDE (2001) *Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection*, University of Dublin.
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.
- Petren K, Grant PR, Grant BR, Clack AA, Lescano NV (2010) Multilocus genotypes from Charles Darwin's finches: biodiversity lost since the voyage of the Beagle. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**, 1009-1018.
- Piry S, Alapetite A, Cornuet JM, et al. (2004) GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity* **95**, 536-539.
- Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity* **90**, 502-503.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
- Prugnolle F, de Meeus T (2002) Inferring sex-biased dispersal from population genetic tools: a review. *Heredity* **88**, 161-165.
- Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. *Evolution* **43**, 258-275.
- Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation Biology* **17**, 230-237.
- Roots EH, Baker RJ (2002) Distribution and characterization of microsatellites in the emu (*Dromaius novaehollandiae*) genome. *Journal of Heredity* **93**, 100-106.

- Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.
- Schwartz MK, Tallmon DA, Luikart G (1998) Review of DNA-based census and effective population size estimators. *Animal Conservation* **1**, 293-299.
- Shaffer ML (1990) Population viability analysis. *Conservation Biology* **4**, 39-40.
- Shepherd LD, Lambert DM (2008) Ancient DNA and conservation: lessons from the endangered kiwi of New Zealand. *Molecular Ecology* **17**, 2174-2184.
- Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15261-15264.
- Taborsky B, Taborsky M (1999) The mating system and stability of pairs in kiwi Apteryx spp. *Journal of Avian Biology* **30**, 143-151.
- Tang B, Huang YH, Lin L, et al. (2003) Isolation and characterization of 70 novel microsatellite markers from ostrich (*Struthio camelus*) genome. *Genome* **46**, 833-840.
- Taylor EL, Blache D, Groth D, Wetherall JD, Martin GB (2000) Genetic evidence for mixed parentage in nests of the emu (*Dromaius novaehollandiae*). *Behavioral Ecology and Sociobiology* **47**, 359-364.
- Taylor EL, Vercoe P, Cockrem J, et al. (1999) Isolation and characterization of microsatellite loci in the emu, *Dromaius novaehollandiae*, and cross-species amplification within Ratitae. *Molecular Ecology* **8**, 1963-1964.
- Turvey ST, Holdaway RN (2005) Postnatal ontogeny, population structure, and extinction of the giant moa Dinornis. *Journal of Morphology* **265**, 70-86.
- van Oosterhout C, Weetman D, Hutchinson WF (2006) Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Molecular Ecology Notes* **6**, 255-256.
- Ward WK, McPartlan HC, Matthews ME, Robinson NA (1998) Ostrich microsatellite polymorphisms at the VIAS-OS4, VIAS-OS8, VIAS-OS14, VIAS-OS22, and VIAS-OS29 loci. *Animal Genetics* **29**, 331-331.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.
- Whitehouse AM, Harley EH (2001) Post-bottleneck genetic diversity of elephant populations in South Africa, revealed using microsatellite analysis. *Molecular Ecology* **10**, 2139-2149.
- Willerslev E, Cooper A (2005) Ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **272**, 3-16.
- Willi Y, Van Buskirk J, Hoffmann AA (2006) Limits to the adaptive potential of small populations. *Annual Review of Ecology Evolution and Systematics* **37**, 433-458.
- Worthy TH, Holdaway RN (1996) Quaternary fossil faunas, overlapping taphonomies, and palaeofaunal reconstruction in north Canterbury, South Island, New Zealand. *Journal of the Royal Society of New Zealand* **26**, 275-361.
- Worthy TH, Holdaway RN (2002) *The lost world of the Moa* Canterbury University Press, Christchurch.
- Wright S (1943) Isolation by distance. *Genetics* **28**, 114-138.

Chapter 6: General discussion

General discussion

This project was founded on DNA from 290 North Canterbury moa individuals of Holocene age. The work has generated 281 Control region Sequences, 227 molecular sex identifications, 217 radiocarbon dates, and 188 multi-locus microsatellite profiles. Having access to 290 fossils excavated from a small confined area, must be characterised as a very rare, if not unique, opportunity in molecular research. Although the dataset covers several thousand years and is therefore not a single genetic snapshot of a population, it is likely the closest we can get to a true population study of extinct species. Moreover, including data from four sympatric taxa and several fossil sites provided some beneficial comparative opportunities for gaining insights on moa biology at several levels.

The previous chapters have discussed aspects of moa biology and evolution, and presented new technical advances to study extinct taxa. However, the moa extinction process has only been assessed briefly in context with the radiocarbon ages in Chapter 4. Before providing a summary of all the main findings, this topic will be revisited here:

A study of moa fossils from natural sites cannot easily be perceived as an extinction study *per se*, because local extinctions went so fast (Holdaway, Jacomb, 2000) that natural deposits are unlikely to reveal any genetic or taphonomic signals of the decline process – except for the fact that no more bones were deposited after 600 BP (see discussion in Chapter 4). The inclusion of material from archaeological sites can potentially be more informative in this regard, and this is assessed in the “Future directions” paragraph below. So rather than dwelling on the extinction, the work presented in this thesis has focused on studying moa populations, as they appeared in a pristine ecosystem before human arrival. Yet, the results are still relevant for the extinction debate. Although the present consensus agrees that humans were largely responsible for the worldwide Pleistocene and early Holocene megafauna extinctions, for example through hunting and inducing an increased frequency of fires (Barnosky et al., 2004; Burney, Flannery, 2005; Gill et al., 2009; Koch, Barnosky, 2006; Prideaux et al., 2007; Roberts, Brook, 2010), it has often been argued that the causes were an interplay between climatic changes and anthropogenic factors (reviewed in Koch, Barnosky, 2006). In regards to the extensive avifauna extinctions in New Zealand (Holdaway, 1989; Tennyson, Martinson, 2006; Worthy, Holdaway, 2002), mid 20th century claims of naturally occurring habitat-changes or “evolutionary dead ends” as causes (reviewed in Worthy, Holdaway, 2002) have now been overturned by overwhelming evidence for hunting, anthropogenic fires and introduction of exotic species such as the pacific rat (*Rattus exulans*) (Anderson, 1989; Holdaway, 1989; Holdaway, 1996; McWethy et al., 2009; Wilmshurst et al., 2008). As for the moa specifically, the consensus is strongly consolidated in favour of a human mediated extinction – most thoroughly advocated by the rapid extinction model in Holdaway and Jacomb (2000). In recent times, only Gemmell et al. (2004) has challenged this contention by suggesting that moa were in decline prior to human arrival, and this publication has

been the impetus for several discussions in the previous chapters. In short, the analyses in Chapters 4 and 5 showed absolute genetic stasis and no loss of genetic diversity throughout mid- to late Holocene. This must be regarded as one of the most important findings, and it is in strong opposition to a pre-human decline hypothesis. In fact, several analyses based on the mtDNA indicate that *Dinornis robustus* (the species that Gemmell et al. (2004) based their predictions on) were more likely to have experienced a Holocene population increase.

However, this point is just one of many new insights presented in the previous chapters. Below are summaries of 12 findings and achievements from this thesis that have all contributed with new knowledge on moa, as well as more technical advances with broader relevance to molecular research. Numbers in brackets refer back to the four main objectives set out in Chapter 1 (Section 2).

Excess of adult females (1)

Of 227 moa individuals that were determined to sex, 190 proved to be females and 37 were males. This is an overall sex ratio across sites and species of 1:5.1 in favour of females. The interpretation is not as straightforward as the result implies, because of depositional biases (see further below). Regardless, it seems obvious that the moa populations displayed a significant excess of females among the adults.

Balanced sex ratios among juveniles (1)

Euryapteryx curtus was represented by a large fraction of juveniles, and it was therefore possible to compare sex ratios in juveniles and adults within this species. The analysis showed a balanced sex ratio among juveniles in contrast to the above mentioned female excess among adults. The ability to extract such detailed information from a taxon that has been extinct for >500 years, is an example of the high level of resolution characterising this project. This result suggested that the excess of females developed in the population as the birds matured, rather than having an embryonic basis. A higher male mortality could perhaps be ascribed to territorial females excluding males from the best habitat and higher stress levels in males, induced by male incubation and paternal parental care as observed among extant ratites (Handford, Mares, 1985).

Compositional differences between sites (1)

Large compositional differences were observed among moa fossils from the two main sites. In terms of sex ratios, an extreme 1:14.2 excess of females was observed at PV whereas a more balanced 1:2.2 was observed at BHV. This highly significant difference can only be explained by a deposition bias at one or both sites. By comparing with previous estimates (Bunce et al., 2003), and by adapting a conservative approach, the 1:2.2 sex ratio at BHV seems like a more realistic reflection of the population than the extreme skew at PV. We hypothesised that the PV deposit “sampled” a disproportionate large fraction of females by being included in the territorial domain for this larger sex. With this argument, the sex ratio topic can be summarised as: (i) a highly significant excess of females was observed. (ii) Balanced sex ratios in juveniles suggested that the overall female excess could partly be ascribed to a higher mortality rate in adult males. (iii) The extreme female excess documented at PV was likely a bias caused by territorial behaviour. The two sites also differed significantly in species composition and fine-scale niche partitioning was suggested as a reasonable explanation – perhaps as an effect of PV and BHV having offered different edge habitats.

The observed compositional differences are very important from a broader paleontological perspective. They emphasize the caution that should be exercised when the recovered material is from a single site only. As seen from the comparative analyses of the PV and BHV fossil record, sampling only one deposit can produce a highly distorted image of the surrounding biodiversity.

The description of a newly discovered fossil site (2)

Although constituting no major taphonomic surprises, the discovery and characterisation of the Rosslea fossil site adds to our knowledge of the pre-human New Zealand environment. The genetic results matched the morphological assessments, but although sufficient to provide a species ID, the DNA preservation in the Rosslea fossils proved disproportionately poor compared to PV and BHV. It is advocated that DNA preservation should be regarded as an important taphonomic character, and that any description of Holocene or Pleistocene fossil sites would benefit from including the molecular element. Moreover, the chapter provides thoroughly outlined sampling protocols, displaying how aDNA technology can work as an integrated part of a multidisciplinary description of a fossil site.

DNA degradation through time (2)

In a comprehensive assessment of ancient DNA degradation through time, it was confirmed that DNA degrades according to an exponential function. This has been hypothesised since the earliest studies by Lindahl (Lindahl, Andersson, 1972; Lindahl, Nyberg, 1972). Although short term observation studies (Bar et al., 1988) and one-point-in-time assessments (Deagle et al., 2006; Schwarz et al., 2009) have suggested that this is indeed the case, this is the first time that it has been demonstrated across a considerable timeframe by the use of ancient DNA – despite many previous attempts.

The South Island moa populations were not panmictic (4)

Phylogenetic analyses of the mtDNA Control region sequences showed that there was distinct geographic structuring in the Holocene, South Island, moa populations. In particular, distinct and basal genetic lineages were present in the North West Nelson area as well as Central Otago. This could indicate that these regions constituted glacial refuges from which moa re-dispersed when the climate and habitat became more favourable. The geographic structure was repeated in several moa species and showed, importantly, that they were not panmictic. This fact seriously questions the method applied by Gemmell et al. (2004), which was used to calculate moa population sizes, leading to the conclusion that moa had declined prior to human arrival in New Zealand. Lastly, the genetic structure documented within the North Canterbury *Dinornis robustus* population could suggest that this area in the late Holocene represented a secondary contact zone of 2-3 old mtDNA lineages.

Barriers to gene flow on the Canterbury Plains (4)

The phylogenetic analyses revealed another interesting feature, in that the genetic connectivity between North and South Canterbury differed among species. *Dinornis robustus* showed no signs of geographic structuring along the East Coast, suggesting that female-mediated gene flow happened freely across the plains. *Euryapteryx curtus* and *Pachyornis elephantopus* on the other hand showed very distinct differentiation between North and South Canterbury, likely facilitated by dispersal barriers. Since the Plains were flat, it seems likely that rivers constituted these barriers and that the large *Dinornis* females were capable of crossing these big rivers, whereas *Euryapteryx* and *Pachyornis* females were not.

The identification and application of microsatellite markers in aDNA research (3)

For the first time, microsatellite markers have been developed directly from aDNA templates of an extinct taxon. This was achieved by using high-throughput sequencing technology, and the method is now used extensively for the identification of microsatellites in modern DNA as well. Six polymorphic microsatellite markers were characterised, and strict protocols were developed to accommodate a high rate of allelic dropout and to ensure sufficient data quality despite the degraded state of aDNA. These advances imply that high resolution microsatellite-based population genetics is now possible in an aDNA context.

The first aDNA microsatellite study (3, 4)

The first aDNA microsatellite study of an extinct taxon was presented in this thesis, including 188 individuals with a multi-locus genetic profile. This was a considerable technical achievement, considering the difficulties in identifying microsatellite markers from aDNA templates, and the subsequent challenges in generating reliable data. The study included a series of microsatellite analyses.

*Very low genetic diversity in *Emeus crassus* (4)*

Both mtDNA and microsatellite data showed that *Emeus crassus* contained considerably less genetic variability than the other species. The result suggested that this species had experienced a major bottleneck at some stage before the mid Holocene – most likely as a result of the last glacial period. *Emeus* was very common on the Eastern Plains in the Holocene, indicating that the population was viable despite exhibiting low genetic diversity through the monitored 3000 years. From a conservation genetic perspective, this is an interesting observation, emphasising that low genetic diversity is not always accompanied by lowered fitness.

*Relatives among the *Dinornis robustus* individuals (4)*

By using the information from several datasets combined (microsatellites, mtDNA and radiocarbon ages), 3-4 pairs of relatives, adult females of *Dinornis*, were identified at Pyramid Valley. The information indicates that females did not disperse far from the area where they hatched. From a more technical perspective, the result is reassuring in that it showed that the microsatellite dataset (in *Dinornis* at least) carries sufficient information to detect such relationships.

Genetic stasis and no indications of a pre-human Holocene decline (4)

A range of methods based on both genetic datasets and all four species were applied in an attempt to uncover any genetic trends throughout the sampled timeframe. The results showed a remarkable temporal stability in the four gene pools. No temporal or spatial genetic structure could be revealed, and individuals separated by thousands of years were not more genetically distant than contemporary individuals. It seems highly unlikely that the moa populations could have maintained the same level of genetic diversity and the same allele frequencies across 3-4000 years, if they were indeed experiencing a major decline as suggested previously (Gemmell et al., 2004). Rather, Bayesian Skyline Plots and starlike haplotype networks indicated that *Dinornis robustus* increased exponentially in population size during the timeframe where it was supposed to be in massive decline. The analyses in Chapters 4 and 5 suggested that the moa populations of North Canterbury were large and viable during the 3000-4000 years leading up to the extinction.

6.1 Future directions

As already outlined in the mini-review in Chapter 1 (Section 1), moa genetics can progress in many interesting directions in the future. Here, I will specifically address some logical extensions of the research presented in the thesis.

The identification of microsatellites from aDNA, and the establishment of viable protocols for generating high quality data, are arguably the most significant technical advances in this project. These achievements have opened the doors for a series of high resolution genetic studies of moa and other extinct taxa. Whereas the analyses presented in Chapter 5 can stand alone as a basic exploration of the first aDNA microsatellite dataset, they also provide the best possible foundation for being followed up with more advanced genetic analyses. As briefly mentioned in Chapter 5, a few clarifying analyses would be beneficial, in order to consolidate the arguments regarding genetic and demographic stability. For example, it must be established that the apparent lack of genetic drift is indeed a demographic feature rather than mirroring limitations in the data. The central question is: Would we be able to detect a demographic bottleneck if it had occurred, given six microsatellite markers with the observed level of genetic variability, and given the very atypical temporal sampling scheme? Once that has been clarified through simulation studies, the next logical development would be an attempt to actually reconstruct the demographic history. This is where the approximate Bayesian computation (ABC) (e.g., Anderson et al., 2005; Cornuet et al., 2008) and/or microsatellite based Skyline Plots (Alexei Drummond, personal communication) may prove essential.

To include samples from other geographic regions would be another obvious expansion of the microsatellite work. For this first exploration of microsatellite data covering thousands of years, it has been advantageous to isolate genetic variability in a temporal dimension, rather than trying to disentangle patterns of genetic differentiation in both space and time. However, now that the scientific viability has been established, it would be highly informative to have comparable data from other regions. This would allow for a detailed study of geographic structure and gene flow among the Holocene moa populations. This could, for example, clarify the strength of the dispersal barriers on the Canterbury Plains. Also from a phylogeographic perspective it would be beneficial to obtain more extensive geographic coverage. The mtDNA analyses showed some very distinct geographic patterns, but the basal lineages were often represented by very few individuals (Figures 5.7a-d). Obtaining mtDNA sequences from more individuals outside of North Canterbury would add significant new insights on the post glacial dispersal and Holocene population structure of moa.

Moa remains from archaeological sites represent a largely untapped resource for aDNA work. By assessing genetics in bones from moa hunter sites, the actual extinction process could be assessed and perhaps shed light on the hunting strategy of the early settlers. For example: Is there any evidence for moa family groups being hunted and killed together? Was there preferential hunting for one of the sexes? If microsatellite data from natural sites and moa hunter sites are analysed together, it may even be possible to assess if hunting only took place in the local area or if large distances were covered on these trips – bringing back locally uncommon genetic profiles. Moreover, eggshell from archaeological sites have proven a great resource for aDNA (Oskam et al., 2010), and molecular work is presently being undertaken to fully explore the potentials with this substrate.

Stable isotopic analyses are becoming increasingly applied in palaeobiology and ecology to study the diet and habitat of extinct and extant organisms (e.g., Hawke, Holdaway, 2005; Iacumin et al., 2000; Palmqvist et al., 2003). The isotopic signatures can also be used to infer environmental conditions such as annual rainfall (Grocke et al., 1997) and average temperatures (Palmqvist et al., 2003) to build up a detailed image of an organism and its environment. Isotopic ratios ($\delta^{15}\text{N}$, $\delta^{13}\text{C}$) in the moa bone collagen are presently being analysed in concert with molecular species and gender identifications to investigate intra- and interspecific niche separation in moa, and to assess the stability of the North Canterbury Holocene environment (Holdaway et al. in prep.). It seems that this coupling of genetic and isotopic signatures provides a very informative framework for studying the biology of extinct species, where observations of behaviour and foraging patterns are non-existent.

Another future project involves aDNA extracted from sediments. This method has previously shown great potential to study ancient environments (Haile et al., 2009; Haile et al., 2007; Matisoo-Smith et al., 2008; Willerslev et al., 2003). In January 2008, a scientific excavation was undertaken at Pyramid Valley with the aim of gaining new insights on the pre-human environment of North Canterbury, and

sediment layers were sampled with sterile equipment (Figure 6.1). Later, aDNA was extracted from these sediments and thousands of chloroplast plant sequences were generated, by applying a combination traditional PCR and subsequent shotgun sequencing of amplicons using GS-FLX (high-throughput) technology. Unfortunately, owing to a lack of reference DNA sequences from New Zealand plants on GenBank, it has not yet been possible to analyse the data in depth, which explains the exclusion from my thesis. However, a small pilot study based on traditional PCR and cloning showed great potential with DNA from numerous plant taxa present (Andersen, 2008). Many reference plants have now been sampled, and this sediment project is soon to regain momentum and will, in time, offer another window into the past.

Figure 6.1: The Pyramid Valley 2008 excavation

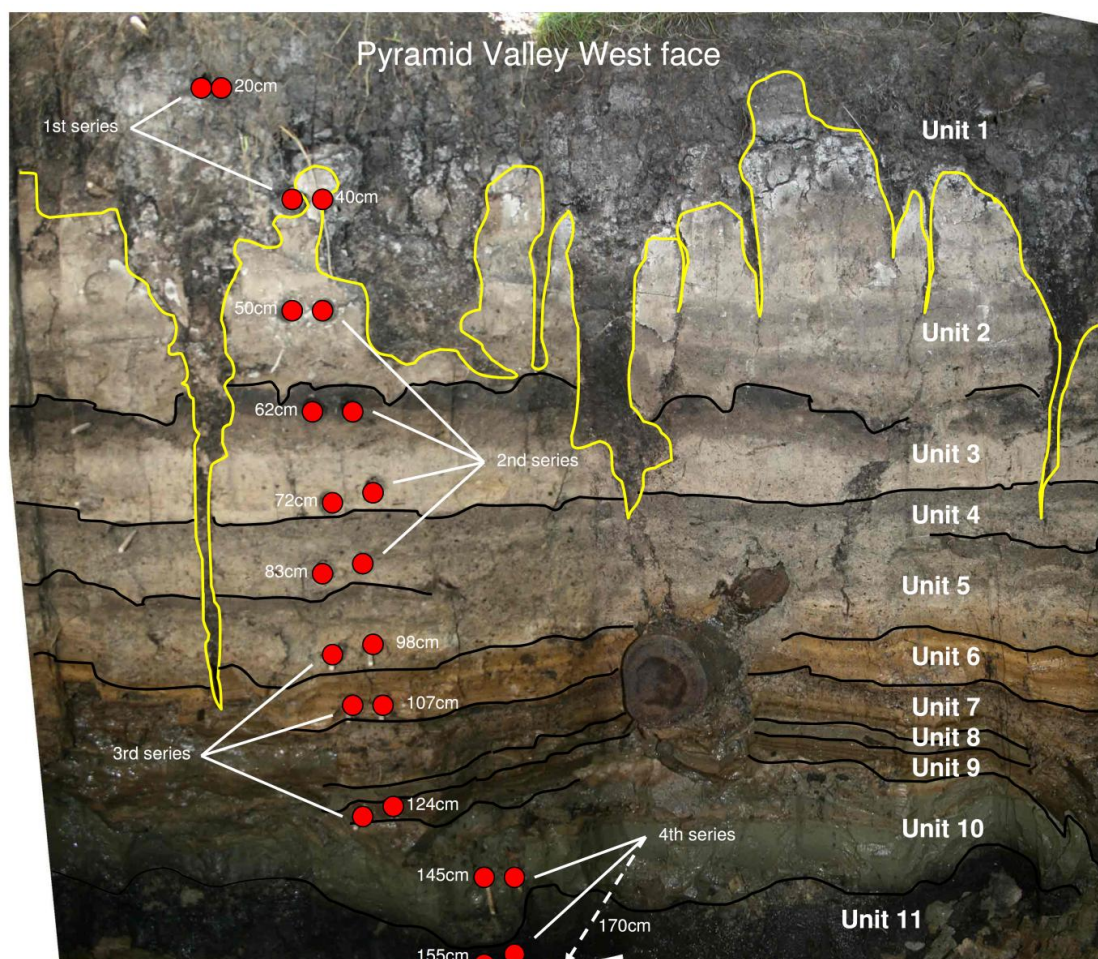


Figure 6.1: A composite image of the western phase as exposed during the Pyramid Valley 2008 excavation. The red marks show where samples were taken for aDNA extraction, to provide a molecular stratigraphy of the fossil site. Other features such as desiccation cracks (yellow lines) and distinct strata (black lines) have been highlighted. A cut matai tree trunk can be seen as well.

6.2 Concluding remarks

The New Zealand moa went extinct >500 years ago, but thanks to aDNA research, an increasingly detailed picture of these iconic giants has been built over the past two decades. The results in this thesis represent significant advances in the field of aDNA as well as in molecular biology in general. The microsatellite study in particular, exemplifies how the research gap between ancient and “modern” genetics is becoming ever smaller. Thanks to the high resolution in these markers, the moa populations have in several aspects been studied almost as if they were still alive.

6.3 References

- Andersen K (2008) aDNA within sediments from a temperate site. Biological Institute, University of Copenhagen.
- Anderson A (1989) *Prodigious Birds* Cambridge University Press, Cambridge, England.
- Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA (2005) Serial SimCoal: A population genetics model for data from multiple populations and points in time. *Bioinformatics* **21**, 1733-1734.
- Bar W, Kratzer A, Machler M, Schmid W (1988) Postmortem stability of DNA. *Forensic Science International* **39**, 59-70.
- Barnosky AD, Koch PL, Feranec RS, Wing SL, Shabel AB (2004) Assessing the causes of Late Pleistocene extinctions on the continents. *Science* **306**, 70-75.
- Bunce M, Worthy TH, Ford T, et al. (2003) Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425**, 172-175.
- Burney DA, Flannery TF (2005) Fifty millennia of catastrophic extinctions after human contact. *Trends in Ecology & Evolution* **20**, 395-401.
- Cornuet JM, Santos F, Beaumont MA, et al. (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**, 2713-2719.
- Deagle BE, Eveson JP, Jarman SN (2006) Quantification of damage in DNA recovered from highly degraded samples--a case study on DNA in faeces. *Front Zool* **3**, 11.
- Gemmell NJ, Schwartz MK, Robertson BC (2004) Moa were many. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, S430-S432.
- Gill JL, Williams JW, Jackson ST, Lininger KB, Robinson GS (2009) Pleistocene Megafaunal Collapse, Novel Plant Communities, and Enhanced Fire Regimes in North America. *Science* **326**, 1100-1103.
- Grocke DR, Bocherens H, Mariotti A (1997) Annual rainfall and nitrogen-isotope correlation in macropod collagen: application as a palaeoprecipitation indicator. *Earth and Planetary Science Letters* **153**, 279-285.
- Haile J, Froese DG, MacPhee RDE, et al. (2009) Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 22352-22357.

- Haile J, Holdaway R, Oliver K, et al. (2007) Ancient DNA chronology within sediment deposits: Are paleobiological reconstructions possible and is DNA leaching a factor? *Molecular Biology and Evolution* **24**, 982-989.
- Handford P, Mares MA (1985) The mating systems of ratites and tinamous - an evolutionary perspective. *Biological Journal of the Linnean Society* **25**, 77-104.
- Hawke DJ, Holdaway RN (2005) Avian assimilation and dispersal of carbon and nitrogen brought ashore by breeding Westland petrels (*Procellaria westlandica*): a stable isotope study. *Journal of Zoology* **266**, 419-426.
- Holdaway RN (1989) New Zealand's pre-human avifauna and its vulnerability. *New Zealand Journal of Ecology* **12**, 11-25.
- Holdaway RN (1996) Arrival of rats in New Zealand. *Nature* **384**, 225-226.
- Holdaway RN, Jacomb C (2000) Rapid extinction of the moas (Aves : Dinorinithiformes): Model, test, and implications. *Science* **287**, 2250-2254.
- Iacumin P, Nikolaev V, Ramigni M (2000) C and N stable isotope measurements on Eurasian fossil mammals, 40 000 to 10 000 years BP: Herbivore physiologies and palaeoenvironmental reconstruction. *Palaeogeography Palaeoclimatology Palaeoecology* **163**, 33-47.
- Koch PL, Barnosky AD (2006) Late quaternary extinctions: State of the debate. *Annual Review of Ecology Evolution and Systematics* **37**, 215-250.
- Lindahl T, Andersson A (1972) Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. *Biochemistry* **11**, 3618-&.
- Lindahl T, Nyberg B (1972) Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **11**, 3610-&.
- Matisoo-Smith E, Roberts K, Welikala N, et al. (2008) Recovery of DNA and pollen from New Zealand lake sediments. *Quaternary International* **184**, 139-149.
- McWethy DB, Whitlock C, Wilmshurst JM, McGlone MS, Li X (2009) Rapid deforestation of South Island, New Zealand, by early Polynesian fires. *Holocene* **19**, 883-897.
- Oskam CL, Haile J, McLay E, et al. (2010) Fossil avian eggshell preserves ancient DNA. *Proc Biol Sci* **277**, 1991-2000.
- Palmqvist P, Grotke DR, Arribas A, Farina RA (2003) Paleoeological reconstruction of a lower Pleistocene large mammal community using biogeochemical ($\delta C-13$, $\delta N-15$, $\delta O-18$, Sr : Zn) and ecomorphological approaches. *Paleobiology* **29**, 205-229.
- Prideaux GJ, Long JA, Ayliffe LK, et al. (2007) An arid-adapted middle Pleistocene vertebrate fauna from south-central Australia. *Nature* **445**, 422-425.
- Roberts RG, Brook BW (2010) Turning back the clock on the extinction of megafauna in Australia. *Quaternary Science Reviews* **29**, 593-595.
- Schwarz C, Debruyne R, Kuch M, et al. (2009) New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Research* **37**, 3215-3229.
- Tennyson A, Martinson P (2006) *Extinct birds of New Zealand* Te Papa Press, Wellington.
- Willerslev E, Hansen AJ, Binladen J, et al. (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300**, 791-795.
- Wilmshurst JM, Anderson AJ, Higham TFG, Worthy TH (2008) Dating the late prehistoric dispersal of polynesians to New Zealand using the commensal Pacific rat. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 7676-7680.
- Worthy TH, Holdaway RN (2002) *The lost world of the Moa* Canterbury University Press, Christchurch.

Appendices

Appendix 1

Contents of the Digital Appendix (DA), included in the accompanying CD:

DA_1a: Alignment (Nexus format) of all *Dinornis robustus* sequences generated in this study

DA_1b: Alignment (Nexus format) of all relevant *Dinornis robustus* sequences available + outgroup

DA_1c: Alignment (Nexus format) of all *Pachyornis elephantopus* sequences generated in this study

DA_1d: Alignment (Nexus format) of all relevant *Pachyornis elephantopus* sequences available + outgroup

DA_1e: Alignment (Nexus format) of all *Euryapteryx curtus* sequences generated on this study

DA_1f: Alignment (Nexus format) of all *Euryapteryx curtus* sequences generated on this study, that also have an associated radiocarbon date

DA_1g: Alignment (Nexus format) of all relevant *Euryapteryx curtus* sequences available + outgroup

DA_1h: Alignment (Nexus format) of all *Emeus crassus* sequences generated on this study

DA_1i: Alignment (Nexus format) of all *Emeus crassus* sequences generated on this study, that also have an associated radiocarbon date

DA_1j: Alignment (Nexus format) of all relevant *Emeus crassus* sequences available + outgroup

DA_2: The ‘master’ database showing relevant information for each sampled fossil

DA_3: Map of the Waikari area in North Canterbury

DA_4: Information and results on all individuals analysed in Chapter 2, Section 1

DA_5: Alignments of the species-specific region, used for species ID in Chapter 2, Section 1

DA_6: Chromatograms showing the principle in moa sex-identification

DA_7: List of samples used in Chapter 3, Section 1

DA_8: Results from qPCR analysis in Chapter 3, Section 1

DA_9: Potential microsatellite markers, detected in the first GS-FLX run

DA_10: Chromatographs showing individual profiles based on the Moa_MS2 marker

DA_11: Allele frequencies relevant to Chapter 3, Section 1

DA_12: Bayesian Skyline Plot of *Dinornis*, assuming a strict molecular clock

DA_13: Overview of the six microsatellite markers, primer sequences, protocols, allele scoring tips etc.

DA_14: Raw microsatellite data (allele calls), representing 188 moa with a multi-locus genotype

DA_15: Expected (H_E) and observed (H_O) heterozygosity for each taxon/locus combination, and other relevant measures

DA_16: Allele frequencies for each of the four taxa and six microsatellites

DA_17: Bar plots of additional STRUCTURE analyses

DA_18: Allentoft et al. (2010): Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand's extinct megafauna. Quaternary Science Reviews, 29, 753-762.

DA_19: Allentoft et al. (2009): Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. Biotechniques, 46, 195-200.

DA_20: Oskam et al. (2010): Fossil avian eggshell preserves ancient DNA. Proc. R. Soc. B. 277, 1991-2000.

DA_21: PhD profile from the Biological Sciences newsletter

DA_22: Podcast from: www.biotechniques.com/multimedia/podcasts/?artId=133170. Interview with M.E. Allentoft and M. Bunce, about 454-sequencing and aDNA (7/4-2009).

DA_23: Abstracts of conference talks

DA_24: Poster presented at conference in Melbourne, 2009: Evolution: The Experience.

Appendix 2

Scientific output during my PhD (3 years):

Publications

Allentoft, M.E.; Bunce, M.; Scofield, R.P.; Hale, M.L.; Holdaway, R.N (2010). Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand's extinct megafauna. *Quaternary Science Reviews*, **29**, 753-762.

Oskam, C.L.; Haile, J.; McLay, E.; Rigby, P.; **Allentoft, M.E.**; Olsen, M.E.; Bengtsson, C.; Miller, G.H.; Schwenninger, J-L.; Jacomb, C.; Walter, R.; Baynes, A.; Dortch, J.; Parker-Pearson, M.; Gilbert, M.T.P.; Holdaway, R.N.; Willerslev, E.; Bunce, M. (2010). Fossil avian eggshell preserves ancient DNA . *Proceedings of the Royal Society B*, **277**, 1991-2000.

Allentoft, M.E.; O'Brien, J. (2010). Global amphibian declines, loss of genetic diversity and fitness: a review. *Diversity*, **2**, 47-71.

Allentoft, M.E.; Siegismund, H.R.; Briggs, L.; Andersen, L.W. (2009). Microsatellite analysis of the natterjack toad (*Bufo calamita*) in Denmark: populations are islands in a fragmented landscape. *Conservation Genetics*, **10**, 15-28.

Allentoft, M.E.; Schuster, S.C.; Holdaway, R.N.; Hale, M.L.; McLay, E.; Oskam, C.; Gilbert, M.T.P.; Spencer, P.; Willerslev, E.; Bunce, M. (2009). Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *BioTechniques*, **46**, 195-200.

Hebsgaard, M.B.; Arneborg, J.; Heyn, P.; **Allentoft, M.E.**; Bunce, M.; Munch, K.; Schweger, C.; Willerslev, E. (2009). The Farm Beneath the Sand – an archaeological case study on ancient ‘dirt’ DNA. *Antiquity*, **83**, 430-444.

Talks:

Allentoft M.E. (presenting), Bunce M., Hale M.L., Holdaway R.N.: Molecules and megafauna: an ancient DNA study of the last moa populations....and other old stuff. Departmental seminar, School of Biological Sciences, University of Canterbury, August, 2010.

Allentoft M.E. (presenting), Holdaway R.N., Hale M.L., Bunce M.: Ancient DNA - new insights into the biology of extinct megafauna. Combined Biological Sciences Meeting (CBSM). September 2009, Perth, Australia.

Allentoft M.E. (presenting), Hale M.L., Bunce M., Holdaway R.N.: Population genetics of the North Canterbury moa. ABCD, School of Biological Sciences, University of Canterbury. 2008.

Poster:

Allentoft, M.E. (presenting), Schuster, S.C., Holdaway, R.N., Hale, M.L., McLay, E., Oskam, C., Gilbert, M.T.P., Spencer, P., Willerslev, E., Bunce, M.: Microsatellites in extinct New Zealand Moa: ancient population genetics goes nuclear. Evolution - The Experience". March 2009, Melbourne.

Appendix 3

Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data

Morten E. Allentoft^{1,4}, Stephan C. Schuster², Richard N. Holdaway^{1,3}, Marie L. Hale¹, Emma McLay⁴, Charlotte Oskam⁴, M. Thomas P. Gilbert⁵, Peter Spencer⁴, Eske Willerslev⁵, and Michael Bunce⁴

¹*School of Biological Sciences, University of Canterbury, Christchurch, New Zealand,* ²*Pennsylvania State University, Center for Comparative Genomics and Bioinformatics, University Park, PA, USA,* ³*Palaecol Research Ltd, Christchurch, New Zealand,* ⁴*Ancient DNA laboratory, School of Biological Sciences and Biotechnology, Murdoch University, Perth, Australia,* and the ⁵*Department of Biology, University of Copenhagen, Copenhagen, Denmark*

BioTechniques 46:195–200 (March 2009) doi 10.2144/000113086

Keywords: ancient DNA; microsatellite development; extinct species; high-throughput 454 sequencing

Supplementary material for this article is available at www.BioTechniques.com/article/113086.

Genetic variation in microsatellites is rarely examined in the field of ancient DNA (aDNA) due to the low quantity of nuclear DNA in the fossil record together with the lack of characterized nuclear markers in extinct species. 454 sequencing platforms provide a new high-throughput technology capable of generating up to 1 gigabases per run as short (200–400-bp) read lengths. 454 data were generated from the fossil bone of an extinct New Zealand moa (Aves: Dinornithiformes). We identified numerous short tandem repeat (STR) motifs, and here present the successful isolation and characterization of one polymorphic microsatellite (Moa_MS2). Primers designed to flank this locus amplified all three moa species tested here. The presented method proved to be a fast and efficient way of identifying microsatellite markers in ancient DNA templates and, depending on biomolecule preservation, has the potential of enabling high-resolution population genetic studies of extinct taxa. As sequence read lengths of the 454 platforms and its competitors (e.g., the SOLEXA and SOLiD platforms) increase, this approach will become increasingly powerful in identifying microsatellites in extinct (and extant) organisms, and will afford new opportunities to study past biodiversity and extinction processes.

Introduction

With the introduction of new high-throughput DNA sequencing techniques capable of generating millions/billions of sequence reads per run, genomic research is advancing faster than ever (1–4). In the field of paleogenetics, the first complete nuclear genome has yet to be recovered, but major sequencing projects of woolly mammoth (5) and Neanderthal (6) are heading what has been called the “third wave” of progress in ancient DNA (aDNA) studies (7). These new sequencing platforms generate large quantities of protein coding data, which will undoubtedly assist in the

study of molecular evolution, functional genomics, and adaptation. In addition to coding data, the huge number of randomly amplified sequences provides the opportunity to search for microsatellites or short tandem repeats (STRs). These non-coding sequences with a high rate of mutation are applied as markers in a wide array of genetic research, especially in relation to forensics, modern population biology, and parentage analyses.

A limited number of successful STR amplifications have been reported from ancient substrates using “modern” STRs as templates (8–10) but to our knowledge, no microsatellite primers developed directly from

aDNA templates (let alone extinct species) have been published prior to this study. This is likely due to complications in traditional microsatellite library construction as the result of the degraded and cross-linked nature of ancient DNA (11). Greenwood et al. (10) managed to amplify a single microsatellite locus in woolly mammoth using primers developed for modern elephants. However, to rely solely on primers developed for related modern taxa, when targeting microsatellites in extinct ones, is problematic because of possible low cross-species amplification rates and chance of monomorphism in the target species (12,13). This is especially pertinent in taxa such as moa, where the likelihood of cross-species amplification is limited as a result of the >80 million years that separate moa from their closest living relatives among the ratite birds (14). Consequently, the chance of identifying polymorphic microsatellite markers in ancient DNA templates of an extinct taxon seems greatly enhanced when the potential markers have been identified directly on the target species.

The Roche GS-FLX (454 Life Sciences, Branford, CT, USA) sequencing technology is currently capable of producing 0.1 gigabases per run with a read length averaging 200–300 nucleotides—a sequence size that allows for the presence of an STR and sufficient flanking regions to design primers. A GS-FLX run was conducted on a *Pachyornis elephantopus* (heavy-footed moa) bone extract to identify a series of microsatellite loci. To illustrate the viability of this technique, we identified an (AC)₁₂ microsatellite (directly from raw GS-FLX data) and demonstrated cross-species amplification in three species of moa. All eleven New Zealand moa species were driven to extinction in the early 15th century, following the arrival of Polynesians. Identification of new STR markers, such as described here, will enable detailed DNA profiling of extinction processes and past population dynamics of these ancient ratites.

Materials and methods

Sampling of moa fossils and DNA isolation

Sampling of moa fossils was conducted by drilling out cylindrical elements (diameter of ~1 cm) of moa tibiotarsal bones using a power drill and diamond-dust coated drill bits. Each sample was then ground into bone powder using a Dremel tool (Part no. 114; Racine, WI, USA). To minimize the incorporation of any possible DNA contamination present on the bones, the bone surfaces and the inner porous parts were excluded, and only solid cortical bone was processed. Contamination from external sources—as well as cross-contami-

nation between samples—was minimized by thoroughly cleaning equipment and sampling environment (with 10% bleach and 100% alcohol) between the processing of each individual. To minimize the risk of a ubiquitous DNA contaminant being present on all the bones, fossils representing three different moa species—from two different sites, and from several different museum/university collections—were included (Supplementary Table 1). Pyramid Valley and Bell Hill Vineyard Swamp both represent late Holocene deposits in North Canterbury, New Zealand, with a known fossil record spanning app. 3700–700 bp (15, R.N.H. unpublished data).

DNA was extracted from 200 mg of bone powder through incubation with rotation at 55°C for 48 h in 1.5 mL digestion buffer [20 mM Tris, pH 8.0, 1% Triton X-100, 10 mM dithiothreitol (DTT), 1 mg/mL proteinase K and 0.5 M EDTA]. The supernatant was spun through Centricon 30,000 molecular weight cut-off (MWCO) ultra-filtration devices (Millipore, Billerica, MA, USA) columns combined with 5 volumes of PBI buffer (Qiagen, Valencia, CA, USA) and DNA was then extracted using silica spin columns (Qiagen) and cleaned with AW1 and AW2 wash buffers (Qiagen) before final elution in 60 µL of 10 mM Tris buffer. For species identification, and to confirm the relative quantities of aDNA in the fossil extracts, a 242-bp sequence of moa mitochondrial control region was amplified (by qPCR) using the species-specific primer set 262f/441r (Supplementary Figure 3) and then sequenced as described in Bunce et al. (16) (data not shown).

GS-FLX sequencing, microsatellite primer development and PCR

A DNA extract from a single moa individual (*P. elephantopus*) from the Bell Hill Vineyard Swamp, which demonstrated good qPCR amplification and long nuclear amplicons (data not shown), was picked for a quarter-plate run on a GS-FLX instrument (Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, PA, USA). The moa DNA library was constructed, as previously described (5,17), with the modification of leaving the extracted DNA unshredded before blunt-ending and phosphorylating the DNA fragments by enzymatic polishing using T4 DNA polymerase, T4 polynucleotide kinase, and *E. coli* DNA polymerase. The blunt-ended, double stranded DNA fragments were then subjected to adapter ligation followed by isolation of the single-stranded template DNA (sstDNA). Subsequently, DNA library fragments were captured onto beads and clonally amplified within individual

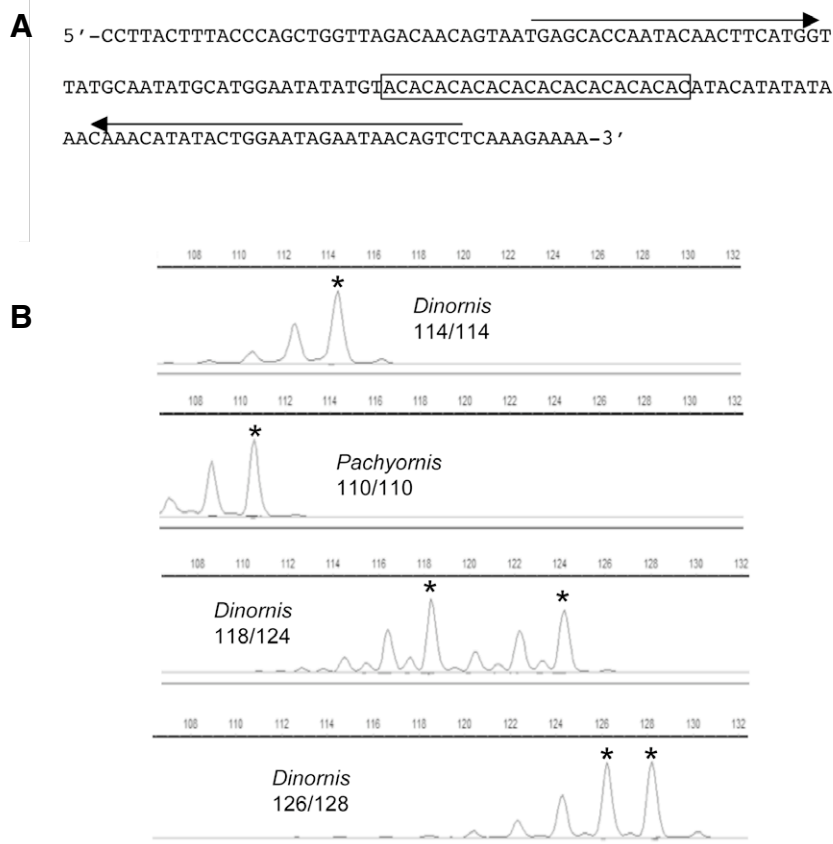


Figure 1. Identification of a polymorphic microsatellite (Moa_MS2) in an extinct species. (A) The original clone identified in the GS-FLX data (clone # 103234_2765_0456). The (AC)₁₂ repetitive region is boxed and the location of forward and reverse primer sequences indicated by arrows. (B) Chromatograms obtained from the amplification of the Moa_MS2 locus in two moa genera (*Dinornis* and *Pachyornis*). Modified screenshot from GENEMARKER version 1.5. The upper two panels show homozygotes and the two lower panels show heterozygotes. Alleles are marked with an asterisk and additional peaks represent stutter and A+ peaks, commonly observed during STR profiling. Chromatograms from other species/individuals are found in Supplementary Figure 1.

emulsion droplets (emPCR). The emulsions were disrupted using isopropanol and beads containing amplified DNA fragments were enriched and recovered for sequencing. The recovered sstDNA beads were packed onto a quarter division of a 70 mm × 75 mm PicoTiterPlate (454 Life Sciences) and loaded onto the GS-FLX sequencing system as previously described (5). The sequencing run yielded 79,796 sequences averaging 112 bp in length. Subsequently, the data were screened for STR sequences using MSATCOMMANDER (18) and a total of 195 di-, tri-, and tetranucleotide repeat sequences were detected (23 of them with six or more repeats). Seven of these proved particularly promising, with flanking regions of sufficient lengths and a base composition suitable for primer design (Figure 1A and Supplementary Figure 2). The sequence chosen for initial trial was clone # 103234_2765_0456, a 158-bp sequence which included an (AC)₁₂ dinucleotide repeat (Figure 1A, GenBank accession no. FJ513189). Primers were then manually designed from this original clone to target

a 114-bp sequence, designated Moa_MS2 (Figure 1A).

Each microsatellite PCR was conducted in a 25-µL volume containing 2 µL DNA extract, 1 µL 10mg/mL bovine serum albumin (BSA), 1 µL each of 10 mM forward (5'-TGAGCACCAATACAACCTTCATGG-3') and reverse primer (5'-GACTGTTATCTATTCCTAGTATATGTTTG-3'), 2.5 µL 10× ABI Gold Buffer, 2.5 µL 25mM MgCl₂, 0.2 µL 5 units/µL ABI Gold Taq polymerase, 0.25 µL of each dNTP (25 mM) and 14.55 µL ultrapure H₂O (Catalog nos. 4311816 and AM 9935; Applied Biosystems, Foster City, CA, USA). Thermal cycling was initiated with a 3-min 95° denaturation step followed by 50 cycles of 95° for 20 s, 57° for 45 s, and 72° for 45 s. A final extension was carried out at 72° for 10 min to maximize adenylation. Negative extraction and amplification controls were always included. DNA fragments were separated on an ABI 3730 genetic analyzer and sized by co-running a Genescan LIZ500 size standard (Applied Biosystems, Foster

Table 1. Summary Statistics for the Moa_MS2 Microsatellite Locus

Species	Size range (bp)	n	N_A	N_E	F_{IS}	Heterozygosity	
						H_O	H_E
South Island giant moa (<i>D. robustus</i>)	110–136	31	10	3.1	0.143	0.581	0.677 ± 0.089
Stout legged moa (<i>E. gravis</i>)	110–128	11	8	5.8	0.230	0.636	0.826 ± 0.145
Heavy footed moa (<i>P. elephantopus</i>)	106–110	10	2	1.9	0.167	0.400	0.505 ± 0.155

Characteristics of the microsatellite locus Moa_MS2 that was optimized for use in three extinct Moa species, including the South Island giant moa (*D. robustus*), stout-legged moa (*E. gravis*) and the heavy-footed moa (*P. elephantopus*). Descriptive measures include n , the number of individuals; N_A , number of alleles; N_E , effective number of alleles; H_O , observed and H_E , expected heterozygosity and F_{IS} , Wright's fixation index as a measure of heterozygote deficiency/excess.

City, CA, USA). DNA fragments were scored manually with the aid of GENEMARKER version 1.5 (Soft Genetics, State College, PA, USA). To check the integrity of the microsatellite, several individual Moa_MS2 amplification products were cloned using a TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA) and sequenced using vector specific M13 primers according to the manufacturer's instructions. Comparisons of the clone sequence data with the original GS-FLX clone confirmed that the Moa_MS2 locus was amplified (data not shown). In accordance with aDNA guidelines, DNA extractions and PCR setup occurred in separate and dedicated aDNA facilities at Murdoch University (Perth, Australia).

Results and discussion

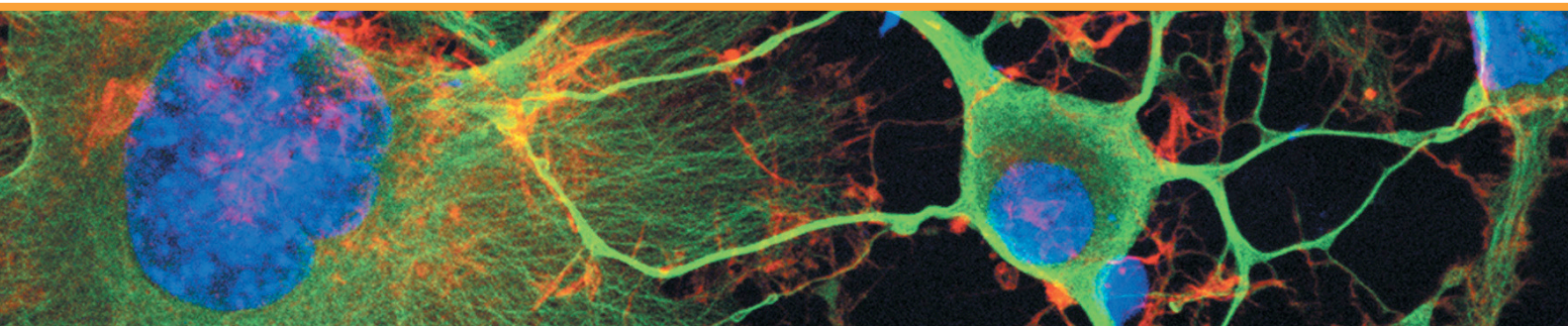
Of the 74 bone DNA extracts prescreened with mtDNA control region amplifications, 52 amplified with the Moa_MS2 primers. The generated chromatographs were consistent with expected dinucleotide STR patterns (including stutters and A+ peaks, see Figure 1B and Supplementary Figure 1). The Moa_MS2 locus yielded a total of 13 different alleles in the three extinct moa genera tested here (*Dinornis*, *Pachyornis*, and *Euryapteryx*). Summary statistics for this single locus are presented in Table 1. Because of the relatively low quality of the template molecules in the fossils, some alleles did not always amplify with the same efficacy in heterozygotes, and allele dropout was observed on occasion. This is not totally unexpected for microsatellites in ancient or

low copy number substrates (19), and when this marker is applied in actual population research (M.A., R.N.H., and M.B., unpublished data), these issues would need to be accounted for through multiple replicates and established protocols (20,21). The chromatographs shown in Figure 1B are a sample of profiles (two homozygotes and two heterozygotes) obtained from 4 moa fossils and demonstrate the 2-bp stuttering characteristic of dinucleotide STR in addition to A+ peaks. Other chromatographs can be found in the supplementary information (Supplementary Figure 1).

The developed primer set amplified across all the three moa genera tested, and—more importantly—displayed polymorphism (Table 1), making it ideal for population studies. Attempts to amplify the Moa_MS2

BioTechniques[®]

The International Journal of Life Science Methods



BioTechniques Weekly

Keep up-to-date with this free weekly electronic newsletter delivered every Thursday to *BioTechniques* subscribers. Coverage includes: **Science news**, **Recent top grants**, **New products**, **Employment opportunities**, **Event calendar**, and more!



Not a subscriber? Sign up free at
www.BioTechniques.com/newsletters

locus in other ratites (emu, ostrich, rhea) were unsuccessful but considering the deep split to the most recent common ancestor, this was not unexpected. Since this study serves to document the use of high-throughput sequence data as a source of microsatellites, a further characterization of the allelic diversity in relation to the various moa species is beyond the scope of this proof-of-concept paper. Clearly, population genetic analyses based on microsatellite data require multiple polymorphic loci. Significantly, our results demonstrate how several potential markers were identified in this quarter-plate GS-FLX run. Although only one STR has been evaluated here, at least seven STRs appeared potentially suitable as genetic markers (Figure S2). Given that birds exhibit a relatively low frequency of microsatellites (22), the frequency of detected microsatellites per run may be even greater in other vertebrate lineages. We screened ~64,000 clones of GenBank mammoth data (5), generated using the GS-20 sequencing platform (454 Life Sciences), for microsatellites and identified 15 STRs with ≥ 6 repeating units (data not shown). The shorter read lengths of the GS-20 (~100 bp), used in the mammoth study, likely explain the slightly lower frequency of identified microsatellites

(0.023%) when compared with the GS-FLX (~200bp) moa run (0.029%).

The characterization of DNA from the fossil record is not without challenges. In fossil bones recovered from soil, a fraction of the present biomolecules is likely to be of bacterial or fungal origin, see (5,23). Moreover, museum specimens are commonly contaminated with human DNA from previous handling (24). When using shotgun sequencing methods such as the ones described here, all DNA molecules in an extract are amplified with equal chance (however, see Reference 25), including any exogenous DNA. Hence, in theory, there is a potential risk of identifying microsatellites—and even conduct population analyses—for the microbial flora/fauna or other contaminants associated with the fossils instead of the target species itself. With regard to the data presented here, we consider human contamination unlikely because the flanking regions of Moa_MS2 demonstrated no significant matches when queried against the human genome database. The paucity of environmental microbe sequences in GenBank makes a similar approach to detect microbes more problematic, but in the case of Moa_MS2, it is unlikely to be of microbial origin for the following reasons:

(i) Though present, microsatellites are much scarcer in prokaryotes and fungus compared with higher eukaryotes (26–29). (ii) Many heterozygotes were detected among the 52 genotyped moa, whereas the haploid nature of bacteria should generate only one allele per clone. (iii) Fungi exist in a variety of ploidy states, and we never observed more than 2 alleles in any of the fossil bone profiles. (iv) Fossils from different collections and originating from very different depositional environments in New Zealand with different microbial communities still yielded reproducible amplifications. (v) Allele frequencies differed among species (see Supplementary Table 2); for example, alleles 106, 114, and 124 are considered “private” alleles for *Pachyornis*, *Dinornis*, and *Euryapteryx* respectively. These findings are incompatible with the possibility that the locus is not of moa origin. (vi) Lastly, the qPCR data presented in Supplementary Figure 3 show a significant difference (*t*-test, $P = 0.000038$) between mean mtDNA control-region Ct values for those DNA extracts that yielded a Moa_MS2 profile (mean Ct = 32.4) and those that yielded mtDNA but failed to amplify the Moa_MS2 locus (mean Ct = 36.7). Fossils with poor mtDNA preservation (high Ct values) had a lower success rate for amplifi-

BioTechniques®

The International Journal of Life Science Methods



Free Subscription

BioTechniques is the first peer-reviewed journal with open access to the whole of the life science community.

- 100% life science
- 100% peer-reviewed methods
- 100% “everyday” practical information

Sign up for free at
www.BioTechniques.com/subscribe

cation of the nuclear marker. A quantitative correlation, such as observed here, should not exist if the microsatellite originated from a source other than moa.

In summary, data generated using a high-throughput DNA sequencing technique were pivotal in identifying a polymorphic microsatellite marker in the extinct New Zealand moa. Mining genomic data for microsatellites is especially relevant to aDNA where the preservation is not conducive to more traditional approaches of library enrichment and construction. We argue that STRs identified using shotgun sequencing approaches, coupled with better DNA isolation techniques, will facilitate research into high-resolution ancient population genetics similar to those currently conducted on extant species.

As average 454 read lengths increase to ~400 bp (30) and the costs of sequencing decrease, the number of potential microsatellite markers identified per run can increase dramatically. It may not be long before shotgun sequencing will replace traditional enrichment library approaches, with one major benefit being the large amount of ancillary mitochondrial and nuclear data that are generated. Moreover, when enrichment steps are combined with coded primers (31), it will be possible to generate microsatellite libraries from multiple species simultaneously. In the context of aDNA, the use of enrichment steps may present a number of additional technical challenges (such as the extent of cross linking) and may be dependent on DNA preservation within the substrate.

Acknowledgements

We would like to gratefully acknowledge the following institutions for samples: Museum of New Zealand Te Papa Tongarewa (A.J.D. Tennyson), Canterbury Museum (P. Scofield) and The Australian Museum of Natural History. We thank Malene Möhl, Fiona Wilson, James Haile, and Ross Barnett for help during the bone sampling process and Lynn Tomsho for technical assistance in the GS-FLX sequencing. Financial support was provided by The Australian Research Council (DP0771971) and the Marsden Fund of the Royal Society of New Zealand (programme 06-PAL-001-EEB) with Palaecol Research, Ltd. Stephan C. Schuster is supported in part by the Gordon and Betty Moore Foundation.

The authors declare no competing interests.

References

- Mardis, E.R. 2008. The impact of next-generation sequencing technology on genetics. *Trends Genet.* 24:133-141.
- Medini, D., D. Serruto, J. Parkhill, D.A. Relman, C. Donati, R. Moxon, S. Falkow, and R. Rappuoli. 2008. Microbiology in the post-genomic era. *Nat. Rev. Microbiol.* 6:419-430.
- Blow, N. 2008. DNA sequencing: generation next-next. *Nat. Methods* 5:267-274.
- Schuster, S.C. 2008. Next-generation sequencing transforms today's biology. *Nat. Methods* 5:16-18.
- Poinar, H.N., C. Schwarz, J. Qi, B. Shapiro, R.D.E. MacPhee, B. Buigues, A. Tikhonov, D.H. Huson, et al. 2006. Metagenomics to paleogenomics: Large-scale sequencing of mammoth DNA. *Science* 311:392-394.
- Green, R.E., J. Krause, S.E. Ptak, A.W. Briggs, M.T. Ronan, J.F. Simons, L. Du, M. Egholm, et al. 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature* 444:330-336.
- Millar, C.D., L. Huynen, S. Subramanian, E. Mohandesan, and D.M. Lambert. 2008. New developments in ancient genomics. *Trends Ecol. Evol.* 23:386-393.
- Amory, S., C. Keyser, E. Crubezy, and B. Ludes. 2007. STR typing of ancient DNA extracted from hair shafts of Siberian mummies. *Forensic Sci. Int.* 166:218-229.
- Manen, J.F., L. Bouby, O. Dalnoki, P. Marinval, M. Turgay, and A. Schlumbaum. 2003. Microsatellites from archaeological *Vitis vinifera* seeds allow a tentative assignment of the geographical origin of ancient cultivars. *J. Archaeol. Sci.* 30:721-729.
- Greenwood, A.D., F. Lee, C. Capelli, R. DeSalle, A. Tikhonov, P.A. Marx, and R.D.E. MacPhee. 2001. Evolution of endogenous retrovirus-like elements of the woolly mammoth (*Mammuthus primigenius*) and its relatives. *Mol. Biol. Evol.* 18:840-847.
- Paabo, S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc. Natl. Acad. Sci. USA* 86:1939-1943.
- Primmer, C.R., A.P. Moller, and H. Ellegren. 1996. A wide-range survey of cross-species microsatellite amplification in birds. *Mol. Ecol.* 5:365-378.
- Primmer, C.R., J.N. Painter, M.T. Koskinen, J.U. Palo, and J. Merila. 2005. Factors affecting avian cross-species microsatellite amplification. *J. Avian Biol.* 36:348-360.
- Cooper, A., C. Lalucza-Fox, S. Anderson, A. Rambaut, J. Austin, and R. Ward. 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* 409:704-707.
- Burrows, C.J. 1989. Moa browsing: evidence from the Pyramid Valley mire. *N. Z. J. Ecol.* 12:51-56.
- Bunce, M., T.H. Worthy, T. Ford, W. Hoppitt, E. Willerslev, A. Drummond, and A. Cooper. 2003. Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* 425:172-175.
- Margulies, M., M. Egholm, W. Altman, S. Attiya, J. Bader, L. Bembgen, J. Berka, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376-380.
- Faircloth, B.C. 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol. Ecol. Resour.* 8:92-94.
- Burger, J., S. Hummel, B. Herrmann, and W. Henke. 1999. DNA preservation: A microsatellite-DNA study on ancient skeletal remains. *Electrophoresis* 20:1722-1728.
- Taberlet, P., S. Griffin, B. Goossens, S. Questiau, V. Manceau, N. Escaravage, L.P. Waits, and J. Bouvet. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res.* 24:3189-3194.
- Morin, P.A., K.E. Chambers, C. Boesch, and L. Vigilant. 2001. Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Mol. Ecol.* 10:1835-1844.
- Primmer, C.R., T. Raudsepp, B.P. Chowdhary, A.R. Moller, and H. Ellegren. 1997. Low frequency of microsatellites in the avian genome. *Genome Res.* 7:471-482.
- Noonan, J.P., M. Hofreiter, D. Smith, J.R. Priest, N. Rohland, G. Rabeder, J. Krause, J.C. Dettler, et al. 2005. Genomic sequencing of Pleistocene cave bears. *Science* 309:597-600.
- Gilbert, M.T.P., L. Rudbeck, E. Willerslev, A.J. Hansen, C. Smith, K.E.H. Penkman, K. Prangenberg, C.M. Nielsen-Marsh, et al. 2005. Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy. *J. Archaeol. Sci.* 32:785-793.
- Dohm, J.C., C. Lottaz, T. Borodina, and H. Himmelbauer. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* 36:e105.
- Karaoglu, H., C.M.Y. Lee, and W. Meyer. 2005. Survey of simple sequence repeats in completed fungal genomes. *Mol. Biol. Evol.* 22:639-649.
- Toth, G., Z. Gaspari, and J. Jurka. 2000. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res.* 10:967-981.
- Schlötterer, C., M. Imhof, H. Wang, V. Nolte, and B. Harr. 2006. Low abundance of *Escherichia coli* microsatellites is associated with an extremely low mutation rate. *J. Evol. Biol.* 19:1671-1676.
- Field, D. and C. Wills. 1998. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. *Proc. Natl. Acad. Sci. USA* 95:1647-1652.
- Droege, M. and B. Hill. 2008. The Genome Sequencer FLX System—longer reads, more applications, straight forward bioinformatics and more complete data sets. *J. Biotechnol.* 136:3-10.
- Binladen, J., M.T.P. Gilbert, J.P. Bollback, F. Panitz, C. Bendixen, R. Nielsen, and E. Willerslev. 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* 2:e197.

Received 15 September 2008; accepted 16 December 2008.

Address correspondence to Michael Bunce, School of Biological Sciences and Biotechnology, Murdoch University, Perth 6150, Australia. email: m.bunce@murdoch.edu.au

Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand's extinct megafauna

Morten E. Allentoft ^{a,b,*}, Michael Bunce ^b, R. Paul Scofield ^c, Marie L. Hale ^a, Richard N.

Holdaway ^{a,d,**}

^a School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

^b Ancient DNA Laboratory, School of Biological Sciences and Biotechnology, Murdoch University, Perth, WA, Australia

^c Canterbury Museum, Rolleston Avenue, Christchurch 8013, New Zealand

^d Palaeocol Research Ltd, P.O. Box 16 569, Hornby, Christchurch 8042, New Zealand

article info

Article history:

Received 30 June 2009

Received in revised form

18 November 2009

Accepted 20 November 2009

abstract

Ancient DNA was isolated from the bones of 267 individuals of the extinct New Zealand moa (Aves: Dinornithiformes) from two late Holocene deposits [Pyramid Valley (PV) and Bell Hill Vineyard (BHV)] located 5.7 km apart in North Canterbury, South Island. The two sites' combined fossil record cover the last 3000 years of pre-human New Zealand and mitochondrial DNA confirmed that four species (*Dinornis robustus*, *Euryapteryx curtus*, *Emeus crassus*, and *Pachyornis elephantopus*) were sympatric in the region. However, the relative species compositions in the two deposits differed significantly with *D. robustus* and *E. crassus* being most abundant at PV while *E. curtus* outnumbered the other three moa taxa combined at BHV. A subsample of 227 individuals had sufficient nuclear DNA preservation to warrant the use of molecular sexing techniques, and the analyses uncovered a remarkable excess of females in both deposits with an overall male to female ratio of 1:5.1. Among juveniles of *E. curtus*, the only species which was represented by a substantial fraction of juveniles, the sex ratio was not skewed (10 ♂, 10 ♀), suggesting that the observed imbalance arose as a result of differential mortality during maturation. Surprisingly, sex ratios proved significantly different between sites with a 1:2.2 ratio at BHV (n = 90) and 1:14.2 at PV (n = 137). Given the mobility of large ratites, and the proximity of the two fossil assemblages in space and time, these differences in taxonomic and gender composition indicate that moa biology and the local environment have affected the fossil representation dramatically and several possible explanations are offered. Apart from adding to our understanding of moa biology, these discoveries reinforce the need for caution when basing interpretation of the fossil record on material from a single site.

© 2009 Elsevier Ltd. All rights reserved.

Doi:10.1016/j.quascirev.2009.11.022

Fossil avian eggshell preserves ancient DNA

Charlotte L. Oskam, James Haile, Emma McLay, Paul Rigby, Morten E. Allentoft, Maia E. Olsen, Camilla Bengtsson, Gifford H. Miller, Jean-Luc Schwenninger, Chris Jacomb, Richard Walter, Alexander Baynes, Joe Dortch, Michael Parker-Pearson, M. Thomas P. Gilbert, Richard N. Holdaway, Eske Willerslev and Michael Bunce

Proc. R. Soc. B published online 10 March 2010
doi: 10.1098/rspb.2009.2019

Supplementary data

["Data Supplement"](#)

<http://rspb.royalsocietypublishing.org/content/suppl/2010/03/09/rspb.2009.2019.DC1.html>

References

[This article cites 37 articles, 12 of which can be accessed free](#)

<http://rspb.royalsocietypublishing.org/content/early/2010/03/09/rspb.2009.2019.full.html#ref-list-1>

P<P

Published online 10 March 2010 in advance of the print journal.

Subject collections

Articles on similar topics can be found in the following collections

[molecular biology](#) (216 articles)

[palaeontology](#) (46 articles)

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *Proc. R. Soc. B* go to: <http://rspb.royalsocietypublishing.org/subscriptions>

Fossil avian eggshell preserves ancient DNA

Charlotte L. Oskam¹, James Haile^{2,3}, Emma McLay¹, Paul Rigby⁴,
Morten E. Allentoft^{1,5}, Maia E. Olsen³, Camilla Bengtsson³,
Gifford H. Miller^{6,7}, Jean-Luc Schwenninger⁸, Chris Jacomb⁹,
Richard Walter⁹, Alexander Baynes¹⁰, Joe Dortch¹¹,
Michael Parker-Pearson¹², M. Thomas P. Gilbert³,
Richard N. Holdaway⁵, Eske Willerslev³ and Michael Bunce^{1,*}

¹Ancient DNA Laboratory, School of Biological Sciences, Murdoch University, Perth, Western Australia 6150, Australia

²Ancient Biomolecules Centre, University of Oxford, Oxford OX1 3PS, UK

³Centre for GeoGenetics, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark

⁴The Centre for Microscopy, Characterization and Analysis, The University of Western Australia, Perth, Western Australia 6009, Australia

⁵School of Biological Sciences, University of Canterbury and Palaecol Research Ltd, Christchurch 8041, New Zealand

⁶Institute of Arctic and Alpine Research (INSTAAR), and ⁷Department of Geological Sciences, University of Colorado, CO 80309-0450, USA

⁸Research Laboratory for Archaeology and the History of Art, University of Oxford, Oxford OX1 3QY, UK

⁹Southern Pacific Archaeological Research, Department of Anthropology, University of Otago, Dunedin 9054, New Zealand

¹⁰Western Australian Museum, Locked Bag 49, Welshpool DC, Western Australia 6986, Australia

¹¹Archaeology, University of Western Australia, Perth, Western Australia 6008, Australia

¹²Department of Archaeology, University of Sheffield, Sheffield S1 4ET, UK

Owing to exceptional biomolecule preservation, fossil avian eggshell has been used extensively in geochronology and palaeodietary studies. Here, we show, to our knowledge, for the first time that fossil eggshell is a previously unrecognized source of ancient DNA (aDNA). We describe the successful isolation and amplification of DNA from fossil eggshell up to 19 ka old. aDNA was successfully characterized from eggshell obtained from New Zealand (extinct moa and ducks), Madagascar (extinct elephant birds) and Australia (emu and owl). Our data demonstrate excellent preservation of the nucleic acids, evidenced by retrieval of both mitochondrial and nuclear DNA from many of the samples. Using confocal microscopy and quantitative PCR, this study critically evaluates approaches to maximize DNA recovery from powdered eggshell. Our quantitative PCR experiments also demonstrate that moa eggshell has approximately 125 times lower bacterial load than bone, making it a highly suitable substrate for high-throughput sequencing approaches. Importantly, the preservation of DNA in Pleistocene eggshell from Australia and Holocene deposits from Madagascar indicates that eggshell is an excellent substrate for the long-term preservation of DNA in warmer climates. The successful recovery of DNA from this substrate has implications in a number of scientific disciplines; most notably archaeology and palaeontology, where genotypes and/or DNA-based species identifications can add significantly to our understanding of diets, environments, past biodiversity and evolutionary processes.

Keywords: ancient DNA; extinct birds; palaeontology; archaeology; eggshell

1. INTRODUCTION

Avian eggshell has been described from numerous fossil deposits and early human settlements all over the world (Miller *et al.* 1999a, 2005; Dortch 2004; Clarke *et al.* 2006). Iconic extinct megafauna, such as the New Zealand moa birds (Aves: Dinornithiformes), the elephant birds of Madagascar (two genera: *Aepyornis* and *Mullerornis*) and Australian thunderbirds (*Genyornis*),

left behind large amounts of this substrate owing to the sheer size and thickness of their eggs. Fossil eggshells have been used extensively to reconstruct palaeoecology and palaeodiets (Miller *et al.* 2005; Clarke *et al.* 2006; Emslie & Patterson 2007), and they also serve as an exceptional medium for a variety of dating methods, including radiocarbon, amino acid racemization and uranium-series disequilibrium (Higham 1994; Johnson *et al.* 1997; Miller *et al.* 1999a; Magee *et al.* 2008).

The matrix of avian eggshell is a highly ordered, porous, biomineral, composed of calcium carbonate (calcite, approx. 97%) and an organic matrix (approx. 3%) (Gautron *et al.* 2001; Nys *et al.* 2004) (figure 1a) and is

* Author for correspondence (m.bunce@murdoch.edu.au).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2009.2019> or via <http://rsob.royalsocietypublishing.org>.

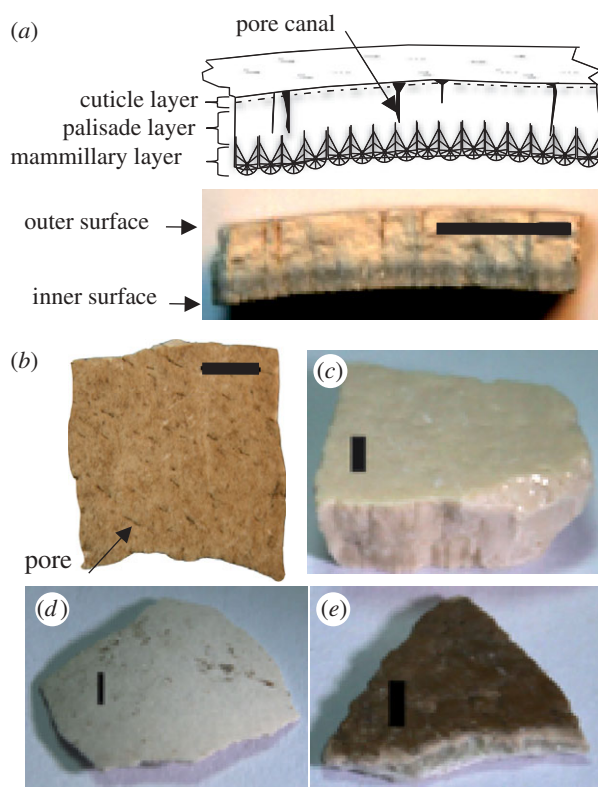


Figure 1. Avian eggshells. (a) Stylized radial cross-section (upper) with corresponding pictorial view (lower) of a moa eggshell (*Dinornis robustus*). (b) Outer surface of a moa eggshell (*D. robustus*). Pores are visible and are aligned towards the poles of the egg. (c) Elephant bird (*Aepyornis maximus*). (d) Duck (*Anas* sp.). (e) Emu (*Dromaius novaehollandiae*). Scale bar, 2 mm.

formed, along with non-calcified membranes, as the egg passes through the oviduct. The eggshell provides the developing avian embryo with an external skeletal support and aids in protection from physical stress and microbes, and controls the exchange of gases and water (Nys *et al.* 2004; Wellman-Labadie *et al.* 2008a). Because most of the organic matrix is intracrystalline, rather than intercrystalline as in most biominerals (Miller *et al.* 2000), avian eggshell suffers little diffusional loss or isotope exchange of its original organic constituents, and the entry of microbes is largely excluded (Miller *et al.* 1992; Johnson & Miller 1997). Despite a documented history of biomolecule preservation, the successful retrieval of DNA from fossil avian eggshell has never been described.

The past decade has seen the field of ancient DNA (aDNA) diversify from more 'traditional' aDNA substrates, such as bone and mummified tissue, into a variety of alternative substrates with varying degrees of success. Some of these include hair and nails; coprolites; sediments (see review, Willerslev & Cooper 2005); feathers (Rawlence *et al.* 2009). Each of these new substrates presents a unique set of challenges; including variation in the levels of contamination from exogenous DNA, types of DNA damage, presence of PCR inhibitors and the absolute quantity of DNA. To complicate matters further, in the age of high-throughput sequencing (HTS), the degree of microbial contamination is as important as the absolute quantity of endogenous DNA in substrate (Gilbert *et al.* 2008). For example, a recent study observed that only

45 per cent of the DNA (sequenced using HTS) obtained from a mammoth bone could be aligned to the elephant genome (Poinar *et al.* 2006). By contrast, Miller *et al.* (2008) aligned approximately 90 and 58 per cent of HTS sequences using DNA isolated from two mammoth hair extractions (Miller *et al.* 2008). Clearly, each substrate has a set of variable preservation characteristics that are dependent on many factors, including the depositional environment.

Recently, DNA was extracted from eggshell membranes and other non-shell residues of bird eggs in museum collections and nest sites (Strausberger & Ashley 2001; Bush *et al.* 2005; Lee & Prys-Jones 2007). However, the only reports of modern DNA being isolated from the eggshell matrix (i.e. without associated membranes) are by Egloff *et al.* (2009) and Rikimaru & Takahashi (2009), who obtained microsatellite DNA profiles from eggshell 'powder' of fresh herring gull (*Larus argentatus*) (Egloff *et al.* 2009) and chicken (*Gallus gallus*) (Rikimaru & Takahashi 2009) eggs. On the basis that amino acids are well preserved in fossil eggshell and given that DNA has been characterized from modern eggshell, we set out to examine the extent of aDNA preservation in fossil eggshell. Furthermore, we critically assess extraction methods and develop a protocol for the efficient isolation and amplification of DNA from this substrate. Our study describes the successful recovery of aDNA from fossil eggshell collected from various archaeological and palaeontological sites in Australia, New Zealand and Madagascar (figure 1 and table 1).

2. MATERIAL AND METHODS

(a) Materials

Initially, we analysed 18 fossil eggshell fragments from 13 locations in Australia, Madagascar and New Zealand, representing a range of climatic conditions. Eggshells were obtained from both museum collections and directly from excavation sites (table 1; table S1a in the electronic supplementary material) and were selected to provide a full assessment of the quality and extent of preservation of DNA in this substrate. Prior to any DNA-based identification, each eggshell was assigned a morphological species identification based on location and/or thickness of the eggshell. For *Dromaius* (emu) and *Genyornis* and the elephant bird eggshells, this was straightforward (Williams 1981), but, because of the range of potential taxa the New Zealand material was unassigned to taxon before DNA sequencing. Where possible, the outermost surfaces of the eggshell were removed by grinding with a Dremel tool (Part no. 114; Racine, WI, USA), to minimize the incorporation of sediments and surface contaminants. Eggshells were then powdered using the same Dremel tool at rotational speeds depending on shell thickness. Powder was collected and weighed in 2.0 ml safelock Eppendorf tubes for later digestion. To prevent contamination from external sources, all samples were prepared in a dedicated aDNA clean room and the sampling area and tools were decontaminated between processing each sample to prevent cross-sample contamination (Cooper & Poinar 2000; Allentoft *et al.* 2009).

(b) Molecular methods

In our initial study, DNA was extracted from eggshell powder using 66–374 mg by rotational incubation at 55°C for

Table 1. A list of bird eggshell specimens from Australia, Madagascar and New Zealand included in this study, with associated primers and GenBank accession numbers. (The following museum abbreviations are used: WRS, Witchcliffe Rock Shelter; MAD, Madagascar; SPAR, Southern Pacific Archaeological Research; WAM, Western Australian Museum; n.a., not attempted; WA, Western Australia; NSW, New South Wales; SI, South Island; NI, North Island.)

site	specimen ID no.	dates	predicted taxa ^a	mtDNA		nuDNA	taxon ^c	common name	GenBank no.
				gene	amplicon size ^b (bp)				
Australia WRS ^d	WAM T20.5.3A	<400	<i>Dromaius</i> sp.	12S ^e , CR ^f	250	yes	<i>D. novaehollandiae</i> (100%)	emu	GU799584–85
	WAM T20.33.6	400–700	<i>Dromaius</i> sp.	12S ^g	125	n.a.	<i>D. novaehollandiae</i> (100%)	emu	GU799586
	WAM G10.5B.21	13 000	<i>Dromaius</i> sp.	12S ^e , CR ^f	250	yes	<i>D. novaehollandiae</i> (100%)	emu	GU799587–88
	WAM G10.106.9	19 000	<i>Dromaius</i> sp.	12S ^g	125	no	<i>D. novaehollandiae</i> (100%)	emu	GU799589
Tallering Hill	WAM_EI	—	unknown	12S ^e	250	yes	<i>Tyto</i> sp. (100%)	owl	GU799590
Popitah ^h	M05-A720	50 000	<i>Genyornis</i>	—	no	n.a.	—	—	—
Eurobilli ^h	M00-A110	50 000	<i>Genyornis</i>	—	no	n.a.	—	—	—
Madagascar Meandre Estuary	MAD 95-47	Holocene	<i>Aepyornis</i> sp.	12S ^g	125	n.a.	<i>M. agilis</i> (92%)	elephant bird	GU799600
	MAD 95-49	Holocene	<i>Aepyornis</i> sp.	12S ^g	125	n.a.	<i>M. agilis</i> (93%)	elephant bird	GU799601
	MAD 97-26	Holocene	<i>Mullerornis</i> sp.	12S ^e	250	n.a.	<i>M. agilis</i> (100%)	elephant bird	GU799591
New Zealand Pyramid Valley, SI	PV_EI	Holocene	unknown	12S ^e	250	yes	<i>Anas</i> sp. (99%)	dabbling duck	GU799592
	SPAR 12	500–700	<i>Dinornis</i> sp.	CR ^j	250	no	<i>D. robustus</i> (100%)	SI giant moa	GU799593
	RL_EI	Holocene	unknown	CR ^j	250	yes	<i>E. crassus</i> (100%)	eastern moa	GU799594
	PN.H18.L.1	400–700	<i>Dinornis</i> sp.	CR ^j	250	yes	<i>D. robustus</i> (99%)	SI giant moa	GU799595
Pounawe ^k , SI	PN.H19.SP2.L1	400–700	<i>Dinornis</i> sp.	CR ^j	250	yes	<i>D. robustus</i> (99%)	SI giant moa	GU799596
	PN.H22.L2	400–700	<i>Dinornis</i> sp.	CR ^j	250	yes	<i>D. robustus</i> (100%)	SI giant moa	GU799597
Hukanui 7a ^l , NI	H7a_E20	> 3000	unknown	CR ^j	250	yes	<i>P. mappini</i> (100%)	heavy footed moa	GU799598
Hukanui Pool ^l , NI	HP_E45	> 3000	unknown	CR ^j	250	yes	<i>A. didiformis</i> (100%)	little bush moa	GU799599

^aBased on eggshell morphology and location.

^bMaximum amplicon length denoted in italics.

^cBased on the closest GenBank BLAST match.

^dDates according to Dortch (2004).

^emtDNA gene region amplified using primers reported in 12Sa/h; Cooper *et al.* (2001).

^fmtDNA gene region amplified using primers reported in emu CR46F/281R present study.

^g12SF5/R4; Haile (2009).

^hDates according to Miller *et al.* (1999b).

ⁱDates according to Jacomb (2009).

^jmoa CR26F/441R; Bunce *et al.* (2003).

^kDates according to Hamel (2001).

^lDates according to Haile *et al.* (2007).

48–72 h in a 750 µl digestion buffer (25 mM Tris pH 8.0 (Sigma, MO, USA), 10 mM dithiothreitol (DTT) (Thermo Fisher Scientific, MA, USA), 1 mg ml⁻¹ proteinase K (Amresco, OH, USA), 0.48 M ethylenediaminetetraacetic acid (EDTA) (Invitrogen, CA, USA) and 0.1 per cent sodium dodecyl sulfate (SDS) (Invitrogen)) or (20 mM Tris, pH 8.0, 10 mM DTT, 1 mg ml⁻¹ proteinase K, 0.47 M EDTA and 1 per cent Triton X-100 (Invitrogen)). After centrifugation, the supernatant was concentrated to approximately 50 µl in a Vivaspin 500 column (MWCO 30000; Sartorius Stedim Biotech, Germany) at 13 000 g, and then combined with five volumes of PBI buffer (Qiagen, CA, USA). DNA was extracted using silica spin columns (Qiagen) and cleaned with 700 µl of AW1 and AW2 wash buffers (Qiagen). Finally, the DNA was eluted from the silica in 60 µl of 10 mM of Tris buffer (Qiagen).

DNA samples were initially screened (by PCR) using a mitochondrial DNA (mtDNA) approximately 250 bp generic bird PCR assay (12Sa/h rRNA) (Cooper *et al.* 2001). Where DNA preservation was poor, shorter fragments were attempted using a variety of species-specific and generic primers (table 1; table S2 in the electronic supplementary material) (Bunce *et al.* 2003). Nuclear DNA (nuDNA) was also assessed, where possible, by amplifying a conserved region of the *c-mos* gene (table S2 in the electronic supplementary material). Multiple negative extraction and amplification controls were included. Selected amplification products were cloned using TOPO TA cloning kits (Invitrogen) and sequenced at a commercial facility, Macrogen (Seoul, South Korea). A subset of samples was sent to Copenhagen and Oxford Universities for independent replication. Sequences were analysed using GENEIOUS 4.8 (Biomatters, New Zealand) and deposited on GenBank (accession numbers GU799584–GU799601).

(c) Optimization of eggshell DNA digestion and extraction methods

Following the successful recovery of DNA from fossil eggshells, different extraction methods were assessed and optimized in order to maximize DNA yields. Eggshells from two moa species in two New Zealand locations, Pounawea (South Island) and Hukanui Pool (North Island), were selected as representative samples. To minimize sample-dependent variation, we used a mixture of eggshell fragments (nine, *Anomalopteryx* sp. from Hukanui Pool and three, *Dinornis* sp. from Pounawea (table S1b in the electronic supplementary material)) that were powdered and pooled to provide a bulk sample suitable for a number of controlled DNA extraction methods. In the first comparison, we tested these bulk samples with three slightly modified extraction methods: (i) chelex with proteinase K (CpK); (ii) phenol/chloroform (P/C); and (iii) digest buffer (DB_a) (see the electronic supplementary material). Variants of these methods are commonly employed on aDNA substrates (Rohland & Hofreiter 2007b; Allentoft *et al.* 2009; Schwarz *et al.* 2009) and recently on modern eggshell (Egloff *et al.* 2009). To standardize the different extraction methods, the following constants were used: 100 mg of eggshell powder, starting digestion volume 750 µl and final DNA elution in 60 µl.

Following the DB_a, chelex and P/C comparisons, we chose the method that generated the highest DNA yield (DB_a method) and proceeded with further optimizations of the protocol. To increase solubility, the DB_a method was

modified by adding a heat step (DB_b) at 95°C, following the initial digestion, this modification was included as default in all subsequent extraction optimization tests. The effect of incubation time was investigated by reducing the 55°C digest from 48 to 2 h (DB_c). Finally, we compared the effect of surfactant types; we tested both anionic surfactants: 1 per cent SDS (DB_d) and non-ionic surfactants: 1 per cent Tween 20 (Sigma) (DB_e), 1 per cent Triton X-100 (DB_f) in addition to the complete absence (DB_g) of surfactants in the DB.

DNA yields were compared using a quantitative PCR (qPCR) assay and a dilution series (2, 0.5 and 0.125 µl) was performed to test for inhibition effects in assay. Furthermore, bovine serum albumin (BSA) was omitted to allow any inhibition effects to be observed. Two qPCR assays (12Sa/h and CR262F/441R) were performed, in duplicate, using the StepOne real-time PCR system (ABI) in 25 µl reactions containing 2 µl DNA extract, 2X Power SYBR Green PCR Master Mix (ABI), 10 µmoles of each primer (table S2 in the electronic supplementary material), and ultrapure H₂O. PCR thermal cycling was initiated with a 10 min 95°C denaturation step followed by 50 cycles of 95°C for 20 s, 58°C for 45 s, followed by an extension at 72°C for 10 min. qPCR data were analysed using the STEPONE 2.0 software. Cycle threshold values were scored at a consistent baseline and values greater than 37, which could not be accurately quantitated, were given a greater than 40 C_T value (tables S3 and S4 in the electronic supplementary material). To determine the relative yield of DNA based on C_T values, two quantitative points (C_T values which differed according to qPCR conventions) were chosen for each method, normalized to provide an estimated C_T value for the undiluted extract and averaged for each PCR assay (see the electronic supplementary material). Student *t*-tests were performed using SPSS 17.0 (IL, USA) to determine whether modifications to the extraction methods generated significantly different DNA yields.

To assess the total DNA preservation, both endogenous and microbial DNA within eggshell, we compared the relative C_T values between avian DNA (12Sa/h) and bacteria (Bac12sqPCR_F and Bac12sqPCR_R) in both eggshell and bone of similar age (table S2 in the electronic supplementary material). Moa bone was isolated according to methods described by Allentoft *et al.* (2009). PCR reactions were set up as for §2c, with the following modifications; the extension step for 16S was altered to a two-step 60°C, (40 cycle) assay for both primer sets and performed in parallel on a MyiQ Real-Time PCR system (Bio-Rad).

(d) Confocal imaging of DNA in fossil eggshells

With the aim of determining the locality of DNA preserved in the eggshell, we employed confocal microscopy together with fluorescent DNA binding dyes. Moa and elephant bird eggshells (figure 2) were incubated at room temperature in Hoechst 33342 (1 mg ml⁻¹, Invitrogen) and 10X SYBR Green (10 000X nucleic acid stain, Invitrogen) dyes for approximately 15 min, then imaged. Hoechst-stained moa eggshell was imaged using the Bio-Rad MRC 1000/10224 ultra violet confocal microscope using a 40×, NA 1.30 oil immersion objective at 488 nm. SYBR-stained elephant bird eggshell was imaged using the Leica SP2 AOBs multi-photon microscope at 5×, NA 0.15 Fluotar objective and excitation at 488 nm. Optical cross-sectional images were taken and analysed in IMAGE J (Rasband 1997–2009).

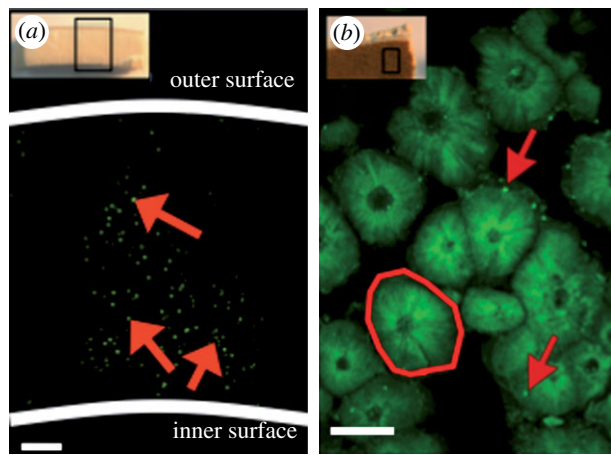


Figure 2. Confocal images of ratite eggshells stained for DNA. (a) Confocal radial cross section of an *Aepyornis* eggshell, stained with SYBR Green, displaying the DNA distributed throughout the matrix: 5 \times objective lens, scale bar, 400 μ m. Inset, orientation of confocal image. (b) Confocal inner surface of a *D. robustus* eggshell, stained with Hoechst dye, displaying mammillary cones (outlined) with peripherally located DNA: 40 \times objective lens, scale bar, 50 μ m. Inset, orientation of confocal image. Red arrows, fluorescently labelled DNA.

3. RESULTS AND DISCUSSION

(a) aDNA is preserved in fossil eggshell

To investigate the level of DNA preservation in fossil eggshell, we sourced samples from a number of palaeontological and archaeological deposits that represent a range of climatic conditions and ages (approx. 400 years to approx. 50 ka old) (table 1; table S1a in the electronic supplementary material). In a set of initial experiments, we attempted to isolate and PCR amplify mtDNA and nuDNA to qualitatively assess biomolecule preservation in these samples. Because there are no existing guidelines on how to maximize DNA recovery from fossil eggshell (see later qPCR data), we employed a methodological approach similar to that used on fossil bone advocated by Rohland & Hofreiter (2007a,b) as the basis for further optimization. We demonstrated that, with the exception of the 50 ka old *Genyornis* eggshell that failed to amplify, we were able to recover DNA from all eggshell sampled. The recovery of aDNA from the eggshell of the extinct *Aepyornis* (figure 1c and table 1) is particularly encouraging, as a number of fossil bones from this genus have failed to yield aDNA. The thickness of *Aepyornis* eggshell readily differentiates it from the smaller *Mullerornis*, and the similarity of both the *Mullerornis* and *Aepyornis* 12S mtDNA sequences to the single published *Mullerornis* GenBank reference sequence (Cooper *et al.* 2001) leaves little doubt as to the authenticity of the sequence. In accordance with aDNA guidelines, identical sequences (table 1) were obtained at both the Murdoch and Oxford aDNA facilities (MAD 95-49) (see the electronic supplementary material) (Haile 2009).

New Zealand has experienced numerous avian extinctions since the arrival of Polynesians ca 700 BP, and the eggshell of the giant ratite moa has been found in many former nesting sites, swamps and archaeological deposits. Table 1 describes the successful characterization of moa DNA from five different deposits of Holocene age, in

both the North and South Islands. The mtDNA sequences recovered from these eggshell samples can be definitively assigned to moa taxa by comparison with fossil bone mtDNA sequences on GenBank. The recovery of *Anas* (dabbling duck, Anatidae) DNA from a tiny piece of thin eggshell from the lake sediments at Pyramid Valley (figure 1d and table 1) demonstrates that DNA preservation is not a feature unique to the thick ratite eggshell; even small starting quantities of eggshell matrix can yield authentic mtDNA and nuDNA. The recovery of sequences tentatively assigned to the genus *Anas* is the first breeding record for that genus from the site (Holdaway & Worthy 1997).

Owing primarily to its hot wet/dry climate, Australia has not traditionally been considered as an environment conducive to long-term DNA preservation (Smith *et al.* 2003). To test if Pleistocene-aged bird eggshell still contains traces of DNA, we investigated material of that age from two Australian species; the extant *Dromaius* (figure 1e) and the extinct *Genyornis* (table 1). We failed to amplify any DNA from *Genyornis* material which is estimated to be approximately 50 ka old (Miller *et al.* 1999b), demonstrating that although eggshell may exhibit properties favouring long-term preservation (see later discussion), it is not exempt from ongoing degradation. By contrast, *Dromaius* eggshell from two limestone cave deposits in southwestern Australia yielded mtDNA and nuDNA sequences from well-dated archaeological contexts up to 19 ka old (table 1) (Dortch 2004). One *Dromaius* eggshell from an archaeological context was charred (possibly as a result of cooking on coals) but still yielded mtDNA (sample AD211; table S1a in the electronic supplementary material). Again, demonstrating that DNA preservation is not restricted to ratite eggshells, we successfully retrieved DNA from a Holocene owl eggshell (Tallering Hill) and identified it as originating from the genus *Tyto* (table 1).

Taken together, the DNA sequences from the samples in table 1 indicate that bird eggshell has the potential for long-term DNA preservation in a number of, often hostile, environments that have not traditionally been conducive to long-term DNA survival. With these encouraging findings, we subsequently went on to investigate first the location of DNA in the eggshell matrix, second the efficiency of various DNA extraction methods and finally the ratio of endogenous avian mtDNA to microbial DNA, in eggshell relative to fossil bone. The purpose of these studies was to provide researchers with some foundation data on how to best approach fossil eggshell from a genetic perspective. Approaches that optimize DNA recovery and enrichment are crucial to this field (Rohland & Hofreiter 2007a,b), for example, it now appears somewhat ironic that aDNA researchers (including authors on this paper) (Cooper *et al.* 2001; Bunce *et al.* 2003) for years, routinely discarded the supernatant following bone powder de-calcification, when this fraction has subsequently been shown to contain a large proportion of the total DNA in the sample (Schwarz *et al.* 2009).

(b) Visualizing aDNA in fossil eggshell

The first step in maximizing DNA recovery from eggshell is to determine where the DNA is physically located. In

modern eggshell, the membranous layers on the inner surface are often used as a source of DNA—given this, the DNA in fossil eggshell might be located solely on the inner surface as the membranes desiccate onto the matrix. If DNA can be identified as concentrated in the inner, outer or calcified layers of eggshell, it would assist in the development of more efficient sampling strategies. Here, we employed two double-stranded DNA binding dyes and confocal imaging techniques to visualize the DNA in this novel substrate (§2).

Confocal imaging of elephant bird (*Aepyornis*) eggshell in cross section clearly demonstrated that DNA is distributed throughout the eggshell matrix as evidenced by the presence of fluorescent ‘hot-spots’ (figure 2*a*). Imaging of the inner surface of moa eggshell (*Dinornis*) also shows foci of DNA clustered around the edges of mammillary cone structures. These images demonstrate that, at least in ratite eggshell, the preserved DNA is distributed somewhat uniformly throughout the eggshell matrix but may be more concentrated around the periphery of the mammillary cones (figure 2*b*). Both the size of the DNA ‘hot-spots’ (2–5 µm) and location, even deep in the eggshell (accessed by optical sectioning), are inconsistent with size and shape of bacteria. Bird eggshell consists of approximately 3 per cent organic matter, composed primarily of intracrystalline proteins (Miller *et al.* 2000) with both structural and antimicrobial functions. One possible hypothesis for the presence of DNA in the eggshell matrix is that when the protein constituents of eggshell are deposited, sloughed epithelial cells are also incorporated into the eggshell calcite (Egloff *et al.* 2009). Given these findings, we see no clear reason to selectively sample specific parts of the eggshell to maximize DNA recovery, other than the obvious precaution of removing debris from the outer surfaces to minimize microbial load. We argue that the entire eggshell matrix can be powdered and undergo DNA isolation. However, in forensic or conservation case studies where non-destructive sampling is required, filing of the outer eggshell surface, as previously demonstrated (Egloff *et al.* 2009), may be more appropriate.

(c) Quantitative and qualitative approaches to optimize aDNA recovery from eggshell

The objective of any DNA extraction technique, especially when dealing with low copy number DNA, is to maximize DNA recovery while at the same time minimizing the co-purification of inhibitors, which can impact on the efficacy of polymerases during PCR. Differences in the chemical composition of the substrates often used in aDNA studies (e.g. bone, hair, sediments) mean that extraction methods need to be modified accordingly to maximize DNA recovery. Most modern biological substrates are so rich in DNA that even if an inappropriate DNA isolation method is employed, some amplifiable DNA is still recovered. This is not a luxury afforded by substrates involved in most aDNA studies, especially if low copy number nuDNA sequences are the targets. The ease or difficulty of aDNA research is linked to two primary factors; DNA preservation and DNA extraction efficacy. With this in mind, we set out to compare a number of commonly employed DNA isolation methods using a similar strategy to Rohland & Hofreiter (2007*a,b*)

Table 2. Results from the qPCR assays comparing three different extraction methods using 100 mg of moa eggshell powder. (CpK, chelex with proteinase K; P/C, phenol chloroform; DB_a, digest buffer with Triton X-100; inhibition column: *, ** and *** inhibition observed at neat, 1/4, and 1/16 concentrations, respectively. Relative to best method values are based on normalized C_T values (tables S3 and S4 in the electronic supplementary material). CpK and P/C methods both performed statistically worse compared with the best performing method (in bold), DB_a (^, $p < 0.001$).)

extraction method	relative to best method	nuDNA		
		80 bp	200 bp	inhibition
CpK	<0.01	yes	no	***
P/C	<0.01	no	no	***
DB _a	1[^]	yes	yes	*

who compared aDNA recovery from Pleistocene-aged cave bear bones and teeth (Rohland & Hofreiter 2007*a,b*).

Here, we compared three methods commonly used in the aDNA/forensic literature; chelex resin (with proteinase K), a P/C method commonly employed in aDNA studies reported in the literature (Cooper *et al.* 2001; Bunce *et al.* 2003) and last, a DB_a in approximately 0.5 M EDTA, which uses a Qiagen silica column (Allentoft *et al.* 2009). A comparison of the mtDNA recovery was made using two independent qPCR assays (§2). In order to observe the extent of PCR inhibition, a serial dilution was performed in the absence of BSA in order to detect inhibition. By pooling eggshell, replicating qPCR assays and using standardized qPCR chemistry, we have ensured the fidelity and reproducibility of the qPCR data presented in tables 2 and 3. Quantitative data presented in table 2 (and tables S3 and S4 in the electronic supplementary material) clearly demonstrate that chelex and P/C methods are uniformly poor in terms of both DNA recovery and the removal of inhibitors. These results may explain why, using modern eggshell, Egloff *et al.* (2009) only achieved 68 per cent success using a P/C method and only 29 per cent using chelex. Both of these methods, in our opinion, are highly unsuitable for DNA recovery and could be one reason why aDNA from fossil bird eggshell had not previously been reported.

Table 3 presents qPCR data representing a further optimization of the extraction protocol. In these experiments, we directly compared three variables: (i) the presence/absence of a heating step at 95°C aimed at increasing the solubility of carbonate salts in eggshell, (ii) digest incubation time, and (iii) the effect of different surfactants. The direct comparisons presented in table 3 demonstrate a number of useful ways to maximize recovery. First, the inclusion of a heat step significantly improved DNA recovery ($p < 0.025$). Second, longer incubation times did not increase DNA yield, in fact approximately 2 h incubations appeared marginally better than longer ones (approx. 48 h), although the difference was not statistically significant. Last, the choice of surfactant also influenced DNA recovery. Tween 20 was always detrimental to recovery ($p < 0.02$), whereas Triton X-100 and SDS generally

Table 3. Results from qPCR assays comparing the initial best performing method from table 2 to the optimized DB extraction methods altering digestion temperature, incubation time and different surfactants using 100 mg of moa eggshell powder. (DB, digest buffer; _a, Triton X-100; _b, 95°C heat step; _c, 2 h incubation; _d, sodium dodecyl sulfate (SDS); _e, Tween 20; _f, without. DB_a performed significantly worse than DB_b (*, $p < 0.025$, paired Students *t*-test). Incubation time had no significant effect on DNA yield (C_T value). Surfactants that performed poorly at a significant level compared with the best performing method (DB_d) are indicated by an asterisk (*, $p < 0.02$ and $p < 0.04$ for DB_e and DB_f, respectively (paired Students *t*-test)). Inhibition column: *, **, inhibition observed at neat and one-quarter concentrations, respectively. Relative to best method values are based on normalized C_T values (tables S3 and S4 in the electronic supplementary material).)

digest buffer optimization		relative to best method			nuDNA		inhibition
		Pounaweia	Hukanui	average	80 bp	200 bp	
best method in table 2	DB _a	0.16	0.52	0.34*	yes	yes	*
temperature effect	DB _b	0.65	0.98	0.81	yes	yes	**
incubation time	DB _c	1	0.88	0.94	yes	yes	**
surfactants							
SDS	DB _d	0.97	1	0.99	yes	yes	**
Tween 20	DB _e	0.19	0.53	0.36*	yes	yes	**
without	DB _f	0.57	0.86	0.72*	yes	yes	**

outperformed those techniques where surfactants were completely omitted ($p < 0.04$). However, the increased viscosity of the DB containing SDS made centrifugation through molecular weight cut-off columns time consuming. Of the methods presented in this study to recover DNA from eggshell, we would recommend the use of a DB containing Triton X-100, an incubation time of between 2 and 24 h, followed by a brief heat step at 95°C. A schematic of our suggested DNA isolation approach for fossil bird eggshell can be found in figure S1, electronic supplementary material.

(d) *Moa eggshell versus bone: qPCR measures of microbial load*

In light of recent advances in DNA sequencing platforms (the Roche GS FLX; ABI SOLiD; Illumina Solexa), and their use in the field of aDNA (Miller *et al.* 2008; Allentoft *et al.* 2009), we set out to compare the relative microbial load in moa bone and moa eggshell. These approaches become increasingly important when deciding upon substrates for de novo whole genome sequencing, and marker discovery, where there is normally little interest in sequencing the vast quantities of uncharacterized plant, fungal and bacterial sequences that might have infiltrated specimens post-mortem. Table 4 presents raw qPCR cycle threshold (C_T) data for a bird-specific assay (12Sa/h) together with a generic 16S assay that amplifies bacterial DNA. By comparing the ΔC_T (the difference in C_T values between bird and bacterial assays) values for six bones and six eggshells, there is more ($p < 0.003$) bacterial DNA present in the moa bone samples compared with moa eggshell. On a cautionary note, although these generic primers have been designed to amplify a wide diversity of bacteria, we cannot rule out biases owing to different depositional environments. The avian to bacterial C_T ratios were on average approximately 125 times (or 6.9 PCR cycles) less in eggshell compared with bone (table 4). Ultimately, relative bacterial loads in eggshell need to be calculated by extensive HTS approaches as have been completed for hair, where favourable ratios resulted in this substrate being chosen for the sequencing of the entire mammoth mitochondrial and nuclear genomes (Gilbert *et al.* 2007;

Miller *et al.* 2008). On the basis of our results, we would suggest that avian eggshell might be a preferable candidate for sequencing extinct avian genomes (mtDNA and nuDNA) including moa, elephant bird, Haast's eagle, the dodo and the great auk, to name a few high profile candidates with interesting evolutionary histories. Alternatively, DNA isolated from feathers might also have similar properties; however, feathers do not preserve as well as eggshell and have to date only been subjected to a limited assessment of aDNA preservation (Rawlence *et al.* 2009).

4. CONCLUSIONS

This study provides, to our knowledge, the first evidence that aDNA is preserved in fossil avian eggshell. Using this substrate, we have managed, for the first time, to our knowledge, to obtain authentic, replicated sequences of DNA from the heaviest bird to have ever existed, the extinct elephant bird *Aepyornis*. The successful retrieval of aDNA from eggshell from Holocene deposits on Madagascar and Pleistocene deposits in Australia is particularly encouraging as neither of these localities have been conducive to long-term DNA preservation in bone. This is most likely owing to a combination of high temperatures and alternating wet/dry cycles. These results raise the question, what properties of eggshell make it conducive to the long-term preservation of biomolecules? We hypothesize that the organic component of eggshell, owing to its stable intracrystalline location, sequesters a small cache of well-preserved DNA as evidenced by the amplification of nuDNA markers (up to 13 ka old) and mtDNA (up to 19 ka old). Bird eggshell is resilient to diffusion losses and acts as a barrier to oxygen and water—the key contributors to aDNA damage via hydrolytic and oxidative mechanisms (Willerslev & Cooper 2005). Interestingly, the organic component of other calcite biominerals, such as mollusc shell, is intercrystalline so different factors may govern biomolecule preservation in these substrates. Modern avian eggshell also exhibits antimicrobial activities (Wellman-Labadie *et al.* 2008a,b) and it is possible that these remain active in fossil eggshell. The low microbe hypothesis is backed up by qPCR data (table 4), which clearly demonstrates a favourable avian

Table 4. A comparison of the relative amounts of moa mtDNA to bacteria DNA in bone and eggshell using qPCR. (Primer sets used in this assay. Samples from Bell Hill and Hukanui are temporarily held by RNH; SPAR, Southern Pacific Archaeological Research.)

specimen ID no.	excavation site	moa taxon	C _T values		ΔC _T [*]	mean ΔC _T
			moa 12S ^a	bacteria 16S ^b		
bone						
BH_B1	Bell Hill	<i>D. robustus</i>	25.64	18.62	7.02	10.38
S39957	Bell Hill	<i>D. robustus</i>	25.58	18.43	7.15	
BH_B2	Bell Hill	<i>P. elephantopus</i>	26.53	17.23	9.3	
SPAR 5	Redcliffs	<i>E. crassus</i>	33.60	16.11	17.50	
SPAR 6	Redcliffs	<i>E. crassus</i>	31.81	19.67	12.14	
SPAR 9	Redcliffs	<i>E. geranoides</i>	26.50	17.31	9.19	
eggshell						
PN.J21.L2	Pounaweia	<i>D. robustus</i>	28.61	26.95	1.66	3.41
PN.H19.SP2.L1	Pounaweia	<i>D. robustus</i>	28.40	26.91	1.49	
PN.H18.L1	Pounaweia	<i>D. robustus</i>	30.53	27.51	3.02	
SPAR 12	Redcliffs	<i>D. robustus</i>	32.54	24.50	8.04	moa : bacteria 1 : 10.6
H7a_E20	Hukanui 7a	<i>P. mappini</i>	28.44	23.27	5.17	
HP_E45	Hukanui Pool	<i>A. didiformis</i>	23.82	22.74	1.08	

* $p < 0.003$ (paired Students *t*-test).

^a12Sa/h; Cooper *et al.* (2001).

^bBac16sqPCR_F/Bac16sqPCR_R (present study) C_T, cycle threshold.

to microbial ratio in moa eggshell when compared with moa bone material of similar age. Excellent mtDNA and nuDNA preservation coupled with low microbial loads make avian eggshell a good candidate for HTS approaches and research is currently underway to identify short tandem repeats markers (Allentoft *et al.* 2009) and single nucleotide polymorphisms for moa and other species of interest.

In an attempt to make this substrate accessible to researchers in the fields of palaeontology and archaeology, we have undertaken a detailed set of quantitative assessments that aim to maximize the recovery of DNA from fossil eggshell. We demonstrate that through the careful selection of isolation methodology and surfactants, aDNA recovery can be maximized. We would like to echo the thoughts of Rohland & Hofreiter (2007b) data on how best to isolate DNA from biological substrates can have tangible effects on the success or failure of aDNA projects. This study serves as a reminder that all non-mineralized substrates will contain biomolecules of some description, in various states of decay and that systematic descriptions of where DNA is located (figure 2) and how to best isolate it (tables 2 and 3) will facilitate genetic analyses of older, more degraded specimens.

The ability to genetically characterize historic and fossil collections of eggs will benefit a number of research programmes including the study of diets and how they have changed over time, and in response to environmental shifts. For example, Antarctic Adélie penguin (*Pygoscelis adeliae*) eggshells up to 38 ka old were recently examined and the $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ isotope values were found to have changed dramatically approximately 200 years ago, prompted by changes in krill availability (Emslie & Patterson 2007). Overlaying both mtDNA and nuDNA genetic signatures (for definitive species and population assignment) together with isotope profiles will assist in data interpretation, particularly in the case of seabird eggshells, which act as a valuable proxy for marine

ecosystem health. Clearly, biomolecules preserved in the matrix of historic and fossil eggshell represent a previously unrecognized and untapped source of DNA, the characterization of which will assist in a range of archaeological, palaeontological, conservation and forensic applications.

We would like to gratefully acknowledge the Western Australian Museum, Canterbury Museum, University of Otago, Palaeocol Research and the Musée d'Art et d'Archéologie of the University of d'Antananarivo for their institutional support. This research was funded by the Australian Research Council (DP0771971 and FT0991741), the Marsden Fund of the Royal Society of New Zealand (09-UOO-164 and 06-PAL-001-EEB) and Murdoch University. For confocal imaging, we acknowledge the support of the Australian Microscopy and Microanalysis Research, UWA, funded by UWA and the state and commonwealth Governments. We would also like to thank the Arts and Humanities Research Council, the British Academy, the British Institute in Eastern Africa, the Danish Natural Science Council, the National Geographic Society, the Natural Environment Research Council, the Nuffield Foundation and the Society of Antiquaries.

REFERENCES

- Allentoft, M. *et al.* 2009 Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *Biotechniques* **46**, 195–200. (doi:10.2144/000113086)
- Bunce, M., Worthy, T. H., Ford, T., Hoppitt, W., Willerslev, E., Drummond, A. & Cooper, A. 2003 Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425**, 172–175. (doi:10.1038/nature01871)
- Bush, K. L., Vinsky, M. D., Aldridge, C. L. & Paszkowski, C. A. 2005 A comparison of sample types varying in invasiveness for use in DNA sex determination in an endangered population of greater sage-grouse (*Centrocercus urophasianus*). *Conserv. Genet.* **6**, 867–870.

- Clarke, S. J., Miller, G. H., Fogel, M. L., Chivas, A. R. & Murray-Wallace, C. V. 2006 The amino acid and stable isotope biogeochemistry of elephant bird (*Aepyornis*) eggshells from southern Madagascar. *Q. Sci. Rev.* **25**, 2343–2356. (doi:10.1016/j.quascirev.2006.02.001)
- Cooper, A. & Poinar, H. N. 2000 Ancient DNA: do it right or not at all. *Science* **289**, 1139. (doi:10.1126/science.289.5482.1139b)
- Cooper, A., Lalueza-Fox, C., Anderson, S., Rambaut, A., Austin, J. & Ward, R. 2001 Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704–707. (doi:10.1038/35055536)
- Dortch, J. 2004 *Palaeo-environmental change and the persistence of human occupation in South-Western Australian forests*. Oxford, UK: Archaeopress.
- Egloff, C., Labrosse, A., Hebert, C. & Crump, D. 2009 A nondestructive method for obtaining maternal DNA from avian eggshells and its application to embryonic viability determination in herring gulls (*Larus argentatus*). *Mol. Ecol. Res.* **9**, 19–27. (doi:10.1111/j.1755-0998.2008.02214.x)
- Emslie, S. D. & Patterson, W. P. 2007 Abrupt recent shift in $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values in Adelie penguin eggshell in Antarctica. *Proc. Natl Acad. Sci. USA* **104**, 11 666–11 669. (doi:10.1073/pnas.0608477104)
- Gautron, J., Hincke, M. T., Mann, K., Panheleux, M., Bain, M., McKee, M. D., Solomon, S. E. & Nys, Y. 2001 Ovocalyxin-32, a novel chicken eggshell matrix protein. *J. Biol. Chem.* **276**, 39 243–39 252. (doi:10.1074/jbc.M104543200)
- Gilbert, M. T. *et al.* 2007 Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* **317**, 1927–1930. (doi:10.1126/science.1146971)
- Gilbert, M. T. P., Miller, W. & Schuster, S. C. 2008 Response to comment on ‘Whole-genome shotgun sequencing of mitochondria from ancient hair shafts’. *Science* **322**, 857b. (doi:10.1126/science.1158565)
- Haile, J. 2009. Ancient DNA from sediments and associated remains. PhD thesis, Department of Zoology, University of Oxford, UK.
- Haile, J. *et al.* 2007 Ancient DNA chronology within sediment deposits: are paleobiological reconstructions possible and is DNA leaching a factor? *Mol. Biol. Evol.* **24**, 982–989. (doi:10.1093/molbev/msm016)
- Hamel, J. 2001 *The archaeology of Otago*. Wellington, New Zealand: Department of Conservation.
- Higham, T. 1994 Radiocarbon dating New Zealand prehistory with moa eggshell: some preliminary results. *Q. Sci. Rev.* **13**, 163–169. (doi:10.1016/0277-3791(94)90043-4)
- Holdaway, R. N. & Worthy, T. H. 1997 A reappraisal of the late Quaternary fossil vertebrates of Pyramid Valley Swamp, North Canterbury, New Zealand. *New Zeal. J. Zool.* **24**, 69–121.
- Jacomb, C. 2009 Excavations and chronology at the Redcliffs Flat site, Canterbury, New Zealand. *Rec. Cant. Mus.* **23**, 15–30.
- Johnson, B. J. & Miller, G. H. 1997 Archaeological applications of amino acid racemization. *Archaeometry* **39**, 265–287. (doi:10.1111/j.1475-4754.1997.tb00806.x)
- Johnson, B. J., Miller, G. H., Fogel, M. L. & Beaumont, P. B. 1997 The determination of the late Quaternary paleoenvironments at Equus Cave, South Africa, using stable isotopes and amino acid racemization in ostrich eggshell. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **136**, 121–137. (doi:10.1016/S0031-0182(97)00043-6)
- Lee, P. L. M. & Prys-Jones, R. P. 2007 Extracting DNA from museum bird eggs, and whole genome amplification of archive DNA. *Mol. Ecol. Notes* **8**, 551–560. (doi:10.1111/j.1471-8286.2007.02042.x)
- Magee, J. W., Miller, G. H., Spooner, N. A., Questiaux, D. G., McCulloch, M. T. & Clark, P. A. 2008 Evaluating Quaternary dating methods: radiocarbon, U-series, luminescence and amino acid racemisation dates of a late Pleistocene emu egg. *Quat. Geochronol.* **4**, 84–92. (doi:10.1016/j.quageo.2008.10.001)
- Miller, G. H., Beaumont, P. B., Jull, A. J. T. & Johnson, B. 1992 Pleistocene geochronology and palaeothermometry from protein diagenesis in ostrich eggshells: implications for the evolution of modern humans. *Phil. Trans. R. Soc. Lond. B* **337**, 149–157. (doi:10.1098/rstb.1992.0092)
- Miller, G. H., Beaumont, P. B., Deacon, H. J., Brooks, A. S., Hare, P. E. & Jull, A. J. T. 1999a Earliest modern humans in southern Africa dated by isoleucine epimerization in ostrich eggshell. *Quat. Sci. Rev.* **18**, 1537–1548. (doi:10.1016/S0277-3791(99)00044-X)
- Miller, G. H., Magee, J. W., Johnson, B. J., Fogel, M. L., Spooner, N. A., McCulloch, M. T. & Ayliffe, L. K. 1999b Pleistocene extinction of *Genyornis newtoni*: human impact on Australian megafauna. *Science* **283**, 205–208. (doi:10.1126/science.283.5399.205)
- Miller, G. H., Hart, C. P., Roark, E. B. & Johnson, B. J. 2000 Isoleucine epimerization in eggshells of the flightless Australian birds, *Genyornis* and *Dromaius*. In *Perspectives in amino acid and protein geochemistry* (eds G. A. Goodfriend, M. J. Collins, M. L. Fogel, S. A. Macko & J. F. Wehmiller), pp. 161–181. Oxford, MA: Oxford University Press.
- Miller, G. H., Fogel, M. L., Magee, J. W., Gagan, M. K., Clarke, S. J. & Johnson, B. J. 2005 Ecosystem collapse in Pleistocene Australia and a human role in megafauna extinction. *Science* **309**, 287–290. (doi:10.1126/science.1111288)
- Miller, W. *et al.* 2008 Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456**, 387–390. (doi:10.1038/nature07446)
- Nys, Y., Gautron, J., Garcia-Ruiz, J. M. & Hincke, M. 2004 Avian eggshell mineralization: biochemical and functional characterization of matrix proteins. *C.R. Palevol.* **3**, 549–562. (doi:10.1016/j.crpv.2004.08.002)
- Poinar, H. N. *et al.* 2006 Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **311**, 392–394. (doi:10.1126/science.1123360)
- Rasband, W. S. 1997–2009 *IMAGE* *®*. Bethesda, MD, USA: US National Institutes of Health. See <http://rsb.info.nih.gov/ij/>.
- Rawlence, N. J., Wood, J. R., Armstrong, K. N. & Cooper, A. 2009 DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa. *Proc. R. Soc. B* **276**, 3395–3402. (doi:10.1098/rspb.2009.0755)
- Rikimaru, K. & Takahashi, H. 2009 A simple and efficient method for extraction of PCR-amplifiable DNA from chicken eggshells. *Anim. Sci. J.* **80**, 220–223. (doi:10.1111/j.1740-0929.2008.00624.x)
- Rohland, N. & Hofreiter, M. 2007a Ancient DNA extraction from bones and teeth. *Nat. Prot.* **2**, 1756–1762. (doi:10.1038/nprot.2007.247)
- Rohland, N. & Hofreiter, M. 2007b Comparison and optimization of ancient DNA extraction. *Biotechniques* **42**, 343–352. (doi:10.2144/000112383)
- Schwarz, C., Debruyne, R., Kuch, M., McNally, E., Schwarcz, H., Aubrey, A. D., Bada, J. & Poinar, H. 2009 New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucl. Acids Res.* **37**, 3215–3229. (doi:10.1093/nar/gkp159)
- Smith, C. I., Chamberlain, A. T., Riley, M. S., Stringer, C. & Collins, M. J. 2003 The thermal history of human fossils and the likelihood of successful DNA amplification.

- J. Hum. Evol.* **45**, 203–217. (doi:10.1016/S0047-2484(03)00106-4)
- Strausberger, B. M. & Ashley, M. V. 2001 Egg yield nuclear DNA from egg-laying female cowbirds, their embryos and offspring. *Conserv. Genet.* **2**, 385–390. (doi:10.1023/A:1012526315617)
- Wellman-Labadie, O., Lakshminarayanan, R. & Hincke, M. T. 2008*a* Antimicrobial properties of avian eggshell-specific C-type lectin-like proteins. *FEBS Lett.* **582**, 699–704. (doi:10.1016/j.febslet.2008.01.043)
- Wellman-Labadie, O., Picman, J. & Hincke, M. 2008*b* Anti-microbial activity of cuticle and outer eggshell protein extracts from three species of domestic birds. *Brit. Poult. Sci.* **49**, 133–143. (doi:10.1016/j.cbpb.2008.01.001)
- Willerslev, E. & Cooper, A. 2005 Ancient DNA. *Proc. R. Soc. B* **272**, 3–16. (doi:10.1098/rspb.2004.2813)
- Williams, D. L. G. 1981 *Genyornis* eggshell (Dromornithidae: Aves) from the Late Pleistocene of South Australia. *Alcheringa* **5**, 133–140.